Electronic Version
Stylesheet Version v1.1.1

Description

BIOINFORMATICALLY DETECTABLE GROUP OF NOVEL REGULATORY OLIGONUCLEOTIDES AND USES THEREOF

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application is a continuation in part of and claims priority from the following patent applications, the disclosures of which applications are all hereby incorporated herein by reference: U.S. Patent Application Serial No. 10/707147 filed 24-Nov-03, U.S. Patent Application Serial No. 10/604985 filed 29-Aug-03, U.S. Patent Application Serial No. 10/651227 filed 29-Aug-03, U.S. Patent Application Serial No. 10/649653 filed 28-Aug-03, U.S. Patent Application Serial No. 10/604926 filed 27-Aug-03, U.S. Patent Application Serial No. 10/604726 filed 13-Aug-03, U.S. Patent Application Serial No. 10/604727 filed 13-Aug-03, and U.S. Provisional Patent Application Serial No. 60/468251 filed 07-May-03. This application also claims priority from International Application Number:

PCT/IL 03/00970, filed 16-Nov-03, the disclosure of which application is hereby incorporated herein by reference. All of the aforesaid patent applications are entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof"; U.S Patent Application Serial No. 10/707147, filed 24-Nov-03, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof" is a continuation in part of and claims priority from the following patent applications, the disclosures of which applications are all hereby incorporated herein by reference: International Application Number: PCT/IL 03/00970, filed 16-Nov-03, U.S. Patent Application Serial No. 10/604985 filed 29-Aug-03, U.S. Patent Application Serial No. 10/651227 filed 29-Aug-03, U.S. Patent Application Serial No. 10/649653 filed 28-Aug-03, U.S. Patent Application Serial No. 10/604926 filed 27-Aug-03, U.S. Patent Application Serial No. 10/604726 filed 13-Aug-03, U.S. Patent Application Serial No. 10/604727 filed 13-Aug-03, and U.S. Provisional Patent Application Serial No. 60/468251 filed 07-May-03. All of the aforesaid patent applications are entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof"; International Application Number: PCT/IL

03/00970, filed 16-Nov-03, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof" is a continuation in part of and claims priority from the following patent applications, the disclosures of which applications are all hereby incorporated herein by reference: U.S. Patent Application Serial No. 10/604985 filed 29-Aug-03, U.S. Patent Application Serial No. 10/651227 filed 29-Aug-03, U.S. Patent Application Serial No. 10/649653 filed 28-Aug-03, U.S. Patent Application Serial No. 10/604926 filed 27-Aug-03, U.S. Patent Application Serial No. 10/604726 filed 13-Aug-03, U.S. Patent Application Serial No. 10/604727 filed 13-Aug-03, U.S. Provisional Patent Application Serial No. 60/468251 filed 07-May-03, and U.S. Patent Application Serial No. 10/345201 filed 16-Jan-03. All of the aforesaid patent applications are entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof"; U.S Patent Application Serial No.10/604985, filed 29-Aug-03, entitled "Bioinformatically is a continuation of U.S Provisional Patent Application Serial No. 60/468251, filed 07-May-03, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof " the disclosure of which is hereby incorporated herein and claims

priority therefrom; and is a continuation in part of and claims priority from the following patent applications, the disclosures of which applications are all hereby incorporated herein by reference: U.S. Patent Application Serial No. 10/651227 filed 29-Aug-03, U.S. Patent Application Serial No. 10/649653 filed 28-Aug-03, U.S. Patent Application Serial No. 10/604926 filed 27-Aug-03, U.S. Patent Application Serial No. 10/604726 filed 13-Aug-03, U.S. Patent Application Serial No. 10/604727 filed 13-Aug-03, U.S. Patent Application Serial No. 10/345201 filed 16-Jan-03, U.S. Patent Application Serial No. 10/321503 filed 18-Dec-02, U.S. Patent Application Serial No. 10/310914 filed 06-Dec-02, and U.S. Patent Application Serial No. 10/293338 filed 14-Nov-02. All of the aforesaid patent applications are entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof"; U.S Patent Application Serial No. 10/604926, filed 27-Aug-03, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof" is a continuation of U.S Patent Application Serial No. 10/345201, filed 16-Jan-03, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof" the disclosure of which is hereby incorporated

herein and claims priority therefrom; and is a continuation in part of and claims priority from the following patent applications, the disclosures of which applications are all hereby incorporated herein by reference:U.S. Patent Application Serial No. 10/604726 filed 13-Aug-03, U.S. Patent Application Serial No. 10/604727 filed 13-Aug-03, U.S. Provisional Patent Application Serial No. 60/468251 filed 07-May-03, U.S. Patent Application Serial No. 10/321503 filed 18-Dec-02, U.S. Patent Application Serial No. 10/310914 filed 06-Dec-02, and U.S. Patent Application Serial No. 10/293338 filed 14-Nov-02. All of the aforesaid patent applications are entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof"; U.S Patent Application Serial No. 10/649653, filed 28-Aug-03, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof" is a continuation of U.S Patent Application Serial No. 10/321503, filed 18-Dec-02, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof "; the disclosure of which is hereby incorporated herein and claims priority therefrom; and is a continuation in part of and claims priority from the following patent applications, the disclosures of which applications are all

hereby incorporated herein by reference: U.S. Patent Application Serial No. 10/604926 filed 27-Aug-03, U.S. Patent Application Serial No. 10/604726 filed 13-Aug-03, U.S. Patent Application Serial No. 10/604727 filed 13-Aug-03, U.S. Provisional Patent Application Serial No. 60/468251 filed 07-May-03, U.S. Patent Application Serial No. 10/321503 filed 18-Dec-02, U.S. Patent Application Serial No. 10/310914 filed 06-Dec-02, and U.S. Patent Application Serial No. 10/293338 filed 14-Nov-02. All of the aforesaid patent applications are entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof"; U.S Patent Application Serial No.10/651227, filed 29-Aug-03, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof" is a continuation of U.S Patent Application Serial No. 10/310914, filed 06-Dec-02, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof "; the disclosure of which is hereby incorporated herein and claims priority therefrom; and is a continuation in part of and claims priority from the following patent applications, the disclosures of which applications are all hereby incorporated herein by reference: U.S. Patent Application Serial No. 10/604985 filed 29-Aug-03,

U.S. Patent Application Serial No. 10/649653 filed 28-Aug-03, U.S. Patent Application Serial No. 10/604926 filed 27-Aug-03, U.S. Patent Application Serial No. 10/604726 filed 13-Aug-03, U.S. Patent Application Serial No. 10/604727 filed 13-Aug-03, U.S. Provisional Patent Application Serial No. 60/468251 filed 07-May-03, U.S. Patent Application Serial No. 10/345201 filed 16-Jan-03, U.S. Patent Application Serial No. 10/321503 filed 18-Dec-02, U.S. Patent Application Serial No. 10/310914 filed 06-Dec-02, and U.S. Patent Application Serial No. 10/293338 filed 14-Nov-02. All of the aforesaid patent applications are entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof"; U.S Patent Applications Serial Nos.10/604727 and 10/604726, filed 13-Aug-03, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof" are a continuation of U.S Patent Application Serial No.10/293338, filed 14-Nov-02, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof ", the disclosure of which is hereby incorporated herein and claims priority therefrom; and are a continuation in part of and claims priority from the following patent applications, the disclosures of which applications

are all hereby incorporated herein by reference: U.S. Provisional Patent Application Serial No. 60/468251 filed 07-May-03, U.S. Patent Application Serial No. 10/345201 filed 16-Jan-03, U.S. Patent Application Serial No. 10/321503 filed 18-Dec-02, U.S. Patent Application Serial No. 10/310914 filed 06-Dec-02, and U.S. Patent Application Serial No. 10/293338 filed 14-Nov-02. All of the aforesaid patent applications are entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof"; U.S Provisional Patent Application Serial No. 60/468251, filed 07-May-03, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof" is a continuation in part of and claims priority from the following patent applications, the disclosures of which applications are all hereby incorporated herein by reference: U.S. Patent Application Serial No. 10/345201 filed 16-Jan-03, U.S. Patent Application Serial No. 10/321503 filed 18-Dec-02, U.S. Patent Application Serial No. 10/310914 filed 06-Dec-02, and U.S. Patent Application Serial No. 10/293338 filed 14-Nov-02. All of the aforesaid patent applications are entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof"; U.S Patent Application Serial No.

10/345201, filed 16-Jan-03, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof" is a continuation in part of and claims priority from the following patent applications, the disclosures of which applications are all hereby incorporated herein by reference: U.S. Patent Application Serial No. 10/321503 filed 18-Dec-02, U.S. Patent Application Serial No. 10/310914 filed 06-Dec-02, and U.S. Patent Application Serial No. 10/293338 filed 14-Nov-02. All of the aforesaid patent applications are entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof"; U.S Patent Application Serial No. 10/321503, filed 18-Dec-02, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof" is a continuation in part of and claims priority from the following patent applications, the disclosures of which applications are all hereby incorporated herein by reference: U.S. Patent Application Serial No. 10/310914 filed 06-Dec-02, and U.S. Patent Application Serial No. 10/293338 filed 14-Nov-02. All of the aforesaid patent applications are entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof"; U.S Patent Application Serial No. 10/310914, filed

06-Dec-02, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof" is a continuation in part of U.S Patent Application Serial No10/293338, filed 14-Nov-02, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof ", the disclosure of which is hereby incorporated by reference and claims priority therefrom.

REFERENCES CITED

[0002] Altschul,S.F., Gish,W., Miller,W., Myers,E.W., and Lipman,D.J. (1990). Basic local alignment search tool. J. Mol. Biol. 215, 403-410.

[0003] Ambros,V., Lee,R.C., Lavanway,A., Williams,P.T., and Jewell,D. (2003). MicroRNAs and Other Tiny Endogenous RNAs in C. elegans 1. Curr. Biol. 13, 807-818.

[0004] Dan Gusfield, Algorithms on strings, trees, and sequences: computer science and computational biology, Cambridge University Press, 1997.

[0005] Elbashir,S.M., Lendeckel,W., and Tuschl,T. (2001). RNA interference is mediated by 21- and 22-nucleotide RNAs. Genes Dev. 15, 188-200.

[0006] Gibbs,W.W. (2003). The unseen genome: gems among the junk. Sci. Am. 289, 46-53.

[0007] Gussow,D. and Clackson,T. (1989). Direct clone character-

ization from plaques and colonies by the polymerase chain reaction. Nucleic Acids Res. 17, 4000.

[0008] Hamosh A, Scott AF, Amberger J, Bocchini C, Valle D and McKusick VA.(2002).Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders.Nucleic Acids Res. 30: 52-55.

[0009] Jenuth,J.P. (2000). The NCBI. Publicly available tools and resources on the Web. Methods Mol. Biol. 132, 301-312.

[0010] Kirkness,E.F. and Kerlavage,A.R. (1997). The TIGR human cDNA database. Methods Mol. Biol. 69, 261-268.

[0011] Lagos-Quintana,M., Rauhut,R., Lendeckel,W., and Tuschl,T. (2001). Identification of novel genes coding for small expressed RNAs. Science 294, 853-858.

[0012] Lau,N.C., Lim,L.P., Weinstein,E.G., and Bartel,D.P. (2001). An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science 294, 858-862.

[0013] Lau,N.C. and Bartel,D.P. (2003). Censors of the genome. Sci. Am. 289, 34-41.

[0014] Lim,L.P., Glasner,M.E., Yekta,S., Burge,C.B., and Bartel,D.P. (2003). Vertebrate microRNA genes. Science 299, 1540.

[0015] Mathews,D.H., Sabina,J., Zuker,M., and Turner,D.H. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary

structure. J. Mol. Biol. 288, 911-940.

[0016] Reinhart,B.J., Slack,F.J., Basson,M., Pasquinelli,A.E., Bettinger,J.C., Rougvie,A.E., Horvitz,H.R., and Ruvkun,G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature 403, 901-906.

[0017] Southern,E.M. (1992). Detection of specific sequences among DNA fragments separated by gel electrophoresis. 1975. Biotechnology 24, 122-139.

[0018] Tom M. Mitchell, Machine Learning, McGraw Hill, 1997.

[0019] Wightman,B., Ha,I., and Ruvkun,G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell 75, 855-862.

[0020] Zhang,H., Kolb,F.A., Brondani,V., Billy,E., and Filipowicz,W. (2002). Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. EMBO J. 21, 5875-5885.

[0021] Zuker,M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31, 3406-3415.

Background of Invention

FIELD OF THE INVENTION

[0022] The present invention relates to a group of bioinformatically detectable novel oligonucleotides, here identified as Genomic Address Messenger or GAM oligonucleotides, which are believed to be related to the micro RNA (miRNA) group of oligonucleotides.

DESCRIPTION OF PRIOR ART

[0023] Micro RNAs (miRNA), are short ~22nt non-coding regulatory RNA oligonucleotides, found in a wide range of species, believed to function as specific gene translation repressors, sometimes involved in cell-differentiation.

[0024] The ability to detect novel miRNAs is limited by the methodologies used to detect such oligonucleotides. All miRNAs identified so far either present a visibly discernable whole body phenotype, as do Lin-4 and Let-7 (Wightman,B., Ha,I., and Ruvkun,G., Cell 75:855-862 (1993); Reinhart et al. Nature 403: 901-906 (2000)), or produce sufficient quantities of RNA so as to be detected by the standard molecular biological techniques.

[0025] Studies reporting miRNAs (Lau et al., Science 294:858-862 (2001), Lagos-Quintana et al., Science 294: 853-858 (2001)) discovered 93 miRNAs in several

species, by sequencing a limited number of clones (300 by Lau and 100 by Lagos-Quintana) of small segments (i.e. size fractionated) RNA. MiRNAs detected in these studies therefore, represent the more prevalent among the miRNA oligonucleotide family, and can not be much rarer than 1% of all small ~20nt-long RNA oligonucleotides.

[0026] The aforesaid studies provide no basis for detection of miRNA oligonucleotides which either do not present a visually discernable whole body phenotype, or are rare (e.g. rarer than 0.1% of all size fractionated ~20nt-long RNA segments expressed in the tissues examined), and therefore do not produce significant enough quantities of RNA so as to be detected by standard biological techniques.

[0027] The following U.S. Patents relate to bioinformatic detection of genes: U.S Patent No. 6,369,195, entitled "Prostate-specific gene for diagnosis, prognosis and management of prostate cancer", and U.S Patent No.6,291,666 entitled "Spike tissue-specific promoter", each of which is hereby incorporated by reference herein.

BRIEF DESCRIPTION OF SEQUENCE LISTING, LARGE TABLES AND COMPUTER PROGRAM LISTING

[0028] A CD including the sequence listing is attached to the

present invention, comprising 142,621 genomic sequences, is contained in a file named SEQ_LIST.TXT (21,399KB, 21-Jan-04), and is hereby incorporated by reference herein.

[0029] A CD including large tables relating to genomic sequences are attached to the present application, appear in 11 table files (size, creation date), incorporated herein: TABLE1.TXT (368 KB, 21-Jan-04); TABLE2.TXT (20,568 KB, 21-Jan-04); TABLE3.TXT (168 KB, 21-Jan-04); TABLE4.TXT (1,191 KB, 21-Jan-04), TABLE5.TXT (152 KB, 21-Jan-04), TABLE6.TXT (9,339 KB, 21-Jan-04) and TABLE7.TXT (197,627 KB, 21-Jan-04), TABLE8.TXT (619,419 KB, 21-Jan-04), TABLE9.TXT (670,091 KB, 21-Jan-04), TABLE10.TXT (1,677 KB, 21-Jan-04) and TABLE11.TXT (54 KB, 21-Jan-04), all of which are incorporated by reference herein.

[0030] A CD including a computer program listing of a computer program constructed and operative in accordance with a preferred embodiment of the present invention is enclosed on an electronic medium in computer readable form, and is hereby incorporated by reference herein The computer program listing is contained in 6 files, the name, sizes and creation date of which are as follows:

AUXILARY_FILES.TXT (117K, 14-Nov-03); BINDING_SITE_SCORING.TXT (17K, 14-Nov-03); EDIT_DISTANCE.TXT (144K,24-Nov-03); FIRST-K.TXT (96K, 24-Nov-03); HAIRPIN_PREDICTION.TXT (47K, 14-Nov-03); TWO_PHASED_SIDE_SELECTOR.TXT (4K, 14-Nov-03); and TWO_PHASED_PREDICTOR.TXT (74K, 14-Nov-03).

Summary of Invention

[0031] The present invention discloses over a thousand novel regulatory microRNA (miRNA) oligonucleotides referred to here as Genomic Address Messenger (GAM) oligonucleotides, which GAM oligonucleotides are detectable using a novel bioinformatic approach, and go undetected by conventional molecular biology methods. Each GAM oligonucleotide specifically inhibits translation of one of more target genes by hybridization of an RNA transcript encoded by the GAM, to a site located in an untranslated region (UTR) of the mRNA of one or more of the target genes.

[0032] The present invention represents a scientific breakthrough, disclosing novel miRNA oligonucleotides the number of which is dramatically larger than previously believed existed. Prior-art studies reporting miRNAs ((Lau et

al., Science 294:858-862 (2001), Lagos-Quintana et al., Science 294: 853-858 (2001)) discovered 93 miRNAs in several species, including 21 in human, using conventional molecular biology methods, such as cloning and sequencing. Bioinformatic detection of miRNA oligonucleotides has not been done prior to the present invention: Despite advanced genome projects, computer-assisted detection of genes encoding functional RNAs remains problematic (Lagos-Quintana et al.,2001).

[0033] Molecular biology methodologies employed by these studies are limited in their ability to detect rare miRNA oligonucleotides, since these studies relied on sequencing of a limited number of clones (300 clones by Lau and 100 clones by Lagos-Quintana) of small segments (i.e. size fractionated) of RNA. MicroRNAs detected in these studies therefore, represent the more prevalent among the miRNA oligonucleotide family, and are typically not be much rarer than 1% of all small ~20nt-long RNA oligonucleotides present in the tissue from the RNA was extracted..

[0034] Recent studies state the number of miRNA genes to be limited, and describe the limited sensitivity of available methods for detection of miRNA: The estimate of 255 human genes is an upper bound implying that no more than

40 miRNA genes remain to be identified in mammals (Lim et al., Science, 299:1540 (2003)); Estimates place the total number of vertebrate miRNA genes at about 200-250 (Ambros et al. Curr. Biol. 13:807-818 (2003)); and Confirmation of very low abundance miRNAs awaits the application of detection methods more sensitive than Northern blots (Ambros, 2003).

[0035] The oligonucleotides of the present invention represent a revolutionary new dimension of genomics and of biology: a dimension comprising a huge number of non-protein coding oligonucleotides which modulate expression of thousands of proteins and are associated with numerous major diseases. This new dimension disclosed by the present invention dismantles a central dogma that has dominated life-sciences during the past 50 years, a dogma which has emphasized the importance of protein coding regions of the genome, holding non-protein coding regions to be of little consequence, often dubbing them junk DNA.

[0036] Indeed, only in recent months has this long held belief as to the low importance of non-protein coding regions been vocally challenged. As an example, an article titled The Unseen Genome - Gems in the Junk (Gibbs,W.W. Sci. Am.

289:46-53 (2003)) asserts that the failure to recognize the importance of non-protein- coding regions may well go down as one of the biggest mistakes in the history of molecular biology. Gibbs further asserts that what was dammned as junk because it was not understood, may in fact turn out to be the very basis of human complexity. The present invention provides a dramatic leap in understanding specific important roles of non-protein coding regions.

[0037] An additional scientific breakthrough of the present invention is a novel conceptual model disclosed by the present invention, which conceptual model is preferably used to encode in a genome the determination of cell-differentiation, utilizing oligonucleotides and polynucleotides of the present invention.

[0038] In various preferred embodiments, the present invention seeks to provide an improved method and system for specific modulation of the expression of specific target genes involved in significant human diseases.It also provides an improved method and system for detection of the expression of novel oligonucleotides of the present invention, which modulate these target genes. In many cases the target genes may be known and fully characterized,

however in alternative embodiments of the present invention, unknown or less well characterized genes may be targeted.

[0039] Accordingly, the invention provides several substantially pure nucleic acids (e.g., genomic DNA, cDNA or synthetic DNA) each comprising a novel GAM oligonucleotide, vectors comprising the DNAs, probes comprising the DNAs, a method and system for selectively modulating translation of known target genes utilizing the vectors, and a method and system utilizing the GAM probes to modulate expression of target genes.

[0040] A Nucleic acid is defined as a ribonucleic acid (RNA) molecule, or a deoxyribonucleic acid (DNA) molecule, or complementary deoxyribonucleic acid (cDNA), comprising either naturally occurring nucleotides or non-naturally occurring nucleotides.

[0041] Substantially pure nucleic acid, Isolated Nucleic Acid, Isolated Oligoucleotide and Isolated Polynucleotide are defined as a nucleic acid that is free of the genome of the organism from which the nucleic acid is derived, and include, for example, a recombinant nucleic acid which is incorporated into a vector, into an autonomously replicating plasmid or virus, or into the genomic nucleic acid of a

prokaryote or eukaryote at a site other than its natural site; or which exists as a separate molecule (e.g., a cDNA or a genomic or cDNA fragment produced by PCR or restriction endonuclease digestion) independent of other nucleic acids.

[0042] An Oligonucleotide is defined as a nucleic acid comprising 2-139 nucleotides, or preferably 16-120 nucleotides. A Polynucleotide is defined as a nucleic acid comprising 140-5000 nucleotides, or preferably 140-1000 nucleotides.

[0043] A Complementary sequence is defined as a first nucleotide sequence which reverse complementary of a second nucleotide sequence: the first nucleotide sequence is reversed relative to a second nucleotide sequence, and wherein each nucleotide in the first nucleotide sequence is complementary to a corresponding nucleotide in the second nucleotide sequence (e.g. ATGGC is the complementary sequence of GCCAT).

[0044] Hybridization, Binding and Annealing are defined as hybridization, under in-vivo physiologic conditions, of a first nucleic acid to a second nucleic acid, which second nucleic acid is at least partially complementary to the first nucleic acid.

[0045] A Hairpin Sturucture is defined as an oligonucleotide having a nucleotide sequence that is 50-140 nucleotides in length, the first half of which nucleotide sequence is at least partially complementary to the second part thereof, thereby causing the nucleic acid to fold onto itself, forming a secondary hairpin structure.

[0046] A Hairpin Shaped Precursor is defined as a Hairpin Structure which is processed by a Dicer enzyme complex, yielding an oligonucleotide which is about 19 to about 24 nucleotides in length.

[0047] "Inhibiting translation" is defined as the ability to prevent synthesis of a specific protein encoded by a respective gene by means of inhibiting the translation of the mRNA of this gene. For example, inhibiting translation may include the following steps: (1) a DNA segment encodes an RNA, the first half of whose sequence is partially complementary to the second half thereof; (2) the precursor folds onto itself forming a hairpin-shaped precursor; (3) a Dicer enzyme complex cuts the hairpin shaped precursor yielding an oligonucleotide that isapproximately 22nt in length; (4) the oligonucleotide binds complementarily to at least one binding site, having a nucleotide sequence that is at least partially complentary to the oligonu-

cleotide, which binding site is located in the mRNA of a target gene, preferably in the untranslated region (UTR) of a target gene, such that the binding inhibts translation of the target protein.

[0048] A "Translation inhibitor site" is defined as the minimal DNA sequence sufficient to inhibit translation.

[0049] There is thus provided in accordance with a preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which is endogenously processed from a hairpin-shaped precursor, and anneals to a portion of a mRNA transcript of a target gene, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs: 1-1436.

[0050] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which is endogenously processed from a hairpin-shaped precursor, and anneals to a portion of a mRNA transcript of a target gene selected from the group consisting of genes shown in Table 11, Row 1, wherein binding of the oligonucleotide

to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs: 1-1436.

[0051] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide having a nucleotide sequence selected from the group consisting of SEQ ID NOs: 1-1436.

[0052] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable first oligonucleotide which is a portion of a mRNA transcript of a target gene, and anneals to a second oligonucleotide that is endogenously processed from a hairpin precursor, wherein binding of the first oligonucleotide to the second oligonucleotide represses expression of the target gene, and wherein nucleotide sequence of the second nucleotide is selected from the group consisting of SEQ ID NOs: 1-1436.

[0053] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable first oligonucleotide which is a portion of a mRNA transcript of a target gene selected from

the group consisting of genes shown in Table 11 row 1, and anneals to a second oligonucleotide that is endogenously processed from a hairpin precursor, wherein binding of the first oligonucleotide to the second oligonucleotide represses expression of the target gene, and wherein nucleotide sequence of the second nucleotide is selected from the group consisting of SEQ ID NOs: 1-1436.

[0054] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable oligonucleotide having a nucleotide sequence selected from the group consisting of SEQ ID NOs: 1-1436.

[0055] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which is endogenously processed from a hairpin-shaped precursor, and anneals to a portion of a mRNA transcript of a target gene associated with Multiple Sclerosis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene.

[0056] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinfor-

matically detectable isolated oligonucleotide which is endogenously processed from a hairpin-shaped precursor, and anneals to a portion of a mRNA transcript of a target gene that is differentially expressed in a tissue affected by Multiple Sclerosis relative an unaffected tissue, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene.

[0057] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which is endogenously processed from a hairpin-shaped precursor, and anneals to a portion of a mRNA transcript of a target gene, the expression of which target gene correlates with Multiple Sclerosis or susceptibility thereto, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene.

[0058] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which is endogenously processed from a hairpin-shaped precursor, and anneals to a portion of a mRNA transcript of a target gene responsible for the formation of Multiple Sclerosis, wherein binding of the oligonucleotide to the mRNA tran-

script represses expression of the target gene.

[0059] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene selected from the group consisting of: CBLB, CNP, CTLA4, HLA-DRA, ICOS, IL12B, IL1RN, ITGA4, MCP, NCF1, NOTCH3, NRG1, PTPRC, PTPRZ1, SPP1 and TNFSF10, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene.

[0060] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which is endogenously processed from a hairpin-shaped precursor, and anneals to a portion of a mRNA transcript of a target gene selected from the group consisting of: CBLB, CNP, CTLA4, HLA-DRA, ICOS, IL12B, IL1RN, ITGA4, MCP, NCF1, NOTCH3, NRG1, PTPRC, PTPRZ1, SPP1 and TNFSF10, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene.

[0061] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which an-

neals to a portion of a mRNA transcript of a target gene selected from the group consisting of: CBLB, CNP, CTLA4, HLA-DRA, ICOS, IL12B, IL1RN, ITGA4, MCP, NCF1, NOTCH3, NRG1, PTPRC, PTPRZ1, SPP1 and TNFSF10, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of: (a) a sequence selected from the group consisting of SEQ ID NOs: 1, 4, 5, 8, 9, 14, 15, 17, 20, 22, 23, 24, 26, 27, 28, 29, 30, 31, 32 ,35 and 4239-4700, and (b) the complement of a sequence selected from the group consisting of SEQ ID NOs: 7889, 7893, 7901, 7918, 7921, 7925, 7946, 8042, 8083, 8089, 8113, 8209, 8258, 8262, 8289, 8304, 8311, 8324, 8377 and 7887-8381.

[0062] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable oligonucleotide having a nucleotide sequence which has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of: (a) a sequence selected from the group consisting of SEQ ID NOs: 1, 4, 5, 8, 9, 14, 15, 17, 20, 22, 23, 24, 26,

27, 28, 29, 30, 31, 32 ,35 and 4239-4700, and (b) the complement of a sequence selected from the group consisting of SEQ ID NOs: 7889, 7893, 7901, 7918, 7921, 7925, 7946, 8042, 8083, 8089, 8113, 8209, 8258, 8262, 8289, 8304, 8311, 8324, 8377 and 7887-8381.

[0063] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Multiple Sclerosis, which target gene is selected from the group consisting of genes shown in Table 11, row 4, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of: (a) a sequence selected from the group consisting of SEQ ID NOs: 1, 4, 5, 8, 9, 14, 15, 17, 20, 22, 23, 24, 26, 27, 28, 29, 30, 31, 32 ,35 and 4239-4700, and (b) the complement of a sequence selected from the group consisting of SEQ ID NOs: 7889, 7893, 7901, 7918, 7921, 7925, 7946, 8042, 8083, 8089, 8113, 8209, 8258, 8262, 8289, 8304, 8311, 8324, 8377 and 7887-8381.

[0064] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene selected from the group consisting of: CBLB, CNP, CTLA4, HLA-DRA, ICOS, IL12B, IL1RN, ITGA4, MCP, NCF1, NOTCH3, NRG1, PTPRC, PTPRZ1, SPP1 and TNFSF10, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs: 1, 4, 5, 8, 9, 14, 15, 17, 20, 22, 23, 24, 26, 27, 28, 29, 30, 31, 32 ,35 and 4239-4700.

[0065] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which includes from about 19 to about 22 nucleotides, which nucleotides are partially complementary to a target gene selected form the group consisting of: CBLB, CNP, CTLA4, HLA-DRA, ICOS, IL12B, IL1RN, ITGA4, MCP, NCF1, NOTCH3, NRG1, PTPRC, PTPRZ1, SPP1 and TNFSF10, and wherein the oligonucleotide is endogenously processed

from a hairpin-form precursor, and includes at least 19 contiguous nucleotides from a sequence selected from the group consisting of SEQ ID NOs: 1, 4, 5, 8, 9, 14, 15, 17, 20, 22, 23, 24, 26, 27, 28, 29, 30, 31, 32 ,35 and 4239-4700.

[0066] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable first oligonucleotide which is a portion of a mRNA transcript of a target gene selected from the group consisting of genes shown in Table 11 row 4, and anneals to a second oligonucleotide that is endogenously processed from a hairpin precursor, wherein binding of the first oligonucleotide to the second oligonucleotide represses expression of the target gene.

[0067] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable first oligonucleotide which is a portion of a mRNA transcript of a target gene selected from the group consisting of genes shown in Table 11 row 4, and anneals to a second oligonucleotide that is endogenously processed from a hairpin precursor, wherein binding of the first oligonucleotide to the second oligonucleotide represses expression of the target gene, and

wherein nucleotide sequence of the second nucleotide is selected from the group consisting of SEQ ID NOs: 1, 4, 5, 8, 9, 14, 15, 17, 20, 22, 23, 24, 26, 27, 28, 29, 30, 31, 32 ,35 and 4239-4700.

[0068] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which is endogenously processed from a hairpin-shaped precursor, and anneals to a portion of a mRNA transcript of a target gene associated with Alzheimers disease, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene.

[0069] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which is endogenously processed from a hairpin-shaped precursor, and anneals to a portion of a mRNA transcript of a target gene that is differentially expressed in a tissue affected by Alzheimers disease relative an unaffected tissue, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene.

[0070] There is still further provided in accordance with another preferred embodiment of the present invention a bioinfor-

matically detectable isolated oligonucleotide which is endogenously processed from a hairpin-shaped precursor, and anneals to a portion of a mRNA transcript of a target gene, the expression of which target gene correlates with Alzheimers disease or susceptibility thereto, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene.

[0071] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which is endogenously processed from a hairpin-shaped precursor, and anneals to a portion of a mRNA transcript of a target gene responsible for the formation of Alzheimers disease, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene.

[0072] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene selected from the group consisting of genes shown in Table 11, row 2, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene.

[0073] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which is endogenously processed from a hairpin-shaped precursor, and anneals to a portion of a mRNA transcript of a target gene selected from the group consisting of genes shown in Table 11, row 2, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene.

[0074] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Alzheimers disease, which target gene is selected from the group consisting of genes shown in Table 11, row 2, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of: (a) a sequence selected from the group consisting of SEQ ID NOs: 1-36 and 1437-4027, and (b) the complement of a sequence selected from the group consisting of SEQ ID NOs: 4999,

5017, 5084, 5137, 5246, 5293, 5319, 5455, 5489, 5538, 5587, 5637, 6039, 6050, 6274, 6334, 6347, 6362, 6417, 6433, 6478, 6575, 6606, 6685, 6809, 6827, 6843, 6869, 6892, 7045, 7321, 7353, 7383, 7442, 7445, 7468 and 4738-7634.

[0075] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene selected from the group consisting of genes shown in Table 11, row 2, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs: 1-36 and 1437-4027.

[0076] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which includes from about 19 to about 22 nucleotides, which nucleotides are partially complementary to a target gene selected form the group consisting of genes shown in Table 11, row 2, and wherein the oligonucleotide is endoge-

nously processed from a hairpin-form precursor, and includes at least 19 contiguous nucleotides from a sequence selected from the group consisting of SEQ ID NOs: 1-36 and 1437-4027.

[0077] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable first oligonucleotide which is a portion of a mRNA transcript of a target gene selected from the group consisting of genes shown in Table 11 row 2, and anneals to a second oligonucleotide that is endogenously processed from a hairpin precursor, wherein binding of the first oligonucleotide to the second oligonucleotide represses expression of the target gene.

[0078] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable first oligonucleotide which is a portion of a mRNA transcript of a target gene selected from the group consisting of genes shown in Table 11 row 2, and anneals to a second oligonucleotide that is endogenously processed from a hairpin precursor, wherein binding of the first oligonucleotide to the second oligonucleotide represses expression of the target gene, and wherein nucleotide sequence of the second nucleotide is

selected from the group consisting of SEQ ID NOs: 1-36 and 1437-4027.

[0079] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which is endogenously processed from a hairpin-shaped precursor, and anneals to a portion of a mRNA transcript of a target gene associated with Duchenne disease, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene.

[0080] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which is endogenously processed from a hairpin-shaped precursor, and anneals to a portion of a mRNA transcript of a target gene that is differentially expressed in a tissue affected by Duchenne disease relative an unaffected tissue, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene.

[0081] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which is endogenously processed from a hairpin-shaped precursor,

and anneals to a portion of a mRNA transcript of a target gene, the expression of which target gene correlates with Duchenne disease or susceptibility thereto, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene.

[0082] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which is endogenously processed from a hairpin-shaped precursor, and anneals to a portion of a mRNA transcript of a target gene responsible for the formation of Duchenne disease, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene.

[0083] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene selected from the group consisting of: CASQ1, DMD, ITGA7, LAMA2, NOS1 and SPARC, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene.

[0084] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinfor-

matically detectable isolated oligonucleotide which is endogenously processed from a hairpin-shaped precursor, and anneals to a portion of a mRNA transcript of a target gene selected from the group consisting of: CASQ1, DMD, ITGA7, LAMA2, NOS1 and SPARC, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene.

[0085] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene selected from the group consisting of: CASQ1, DMD, ITGA7, LAMA2, NOS1 and SPARC, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of: (a) a sequence selected from the group consisting of SEQ ID NOs: 5, 8, 11, 14, 18, 24, 26, 27, 28, 29, 30, 33, 35, 36 and 4028-4238, and (b) the complement of a sequence selected from the group consisting of SEQ ID NOs: 7644, 7672, 7697, 7737, 7739, 7755, 7766, 7768, 7780, 7791, 7793, 7809, 7815, 7868 and 7635-7886.

[0086] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Duchenne disease, which target gene is selected from the group consisting of genes shown in Table 11, row 3, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of: (a) a sequence selected from the group consisting of SEQ ID NOs: 5, 8, 11, 14, 18, 24, 26, 27, 28, 29, 30, 33, 35, 36 and 4028-4238, and (b) the complement of a sequence selected from the group consisting of SEQ ID NOs: 7644, 7672, 7697, 7737, 7739, 7755, 7766, 7768, 7780, 7791, 7793, 7809, 7815, 7868 and 7635-7886.

[0087] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene selected from the group consisting of: CASQ1, DMD, ITGA7, LAMA2, NOS1 and SPARC, wherein binding of the

oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs: 5, 8, 11, 14, 18, 24, 26, 27, 28, 29, 30, 33, 35, 36 and 4028-4238.

[0088] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which includes from about 19 to about 22 nucleotides, which nucleotides are partially complementary to a target gene selected form the group consisting of: CASQ1, DMD, ITGA7, LAMA2, NOS1 and SPARC, and wherein the oligonucleotide is endogenously processed from a hairpin-form precursor, and includes at least 19 contiguous nucleotides from a sequence selected from the group consisting of SEQ ID NOs: 5, 8, 11, 14, 18, 24, 26, 27, 28, 29, 30, 33, 35, 36 and 4028-4238.

[0089] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable first oligonucleotide which is a portion of a mRNA transcript of a target gene selected from the group consisting of genes shown in Table 11 row 3,

and anneals to a second oligonucleotide that is endogenously processed from a hairpin precursor, wherein binding of the first oligonucleotide to the second oligonucleotide represses expression of the target gene.

[0090] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable first oligonucleotide which is a portion of a mRNA transcript of a target gene selected from the group consisting of genes shown in Table 11 row 3, and anneals to a second oligonucleotide that is endogenously processed from a hairpin precursor, wherein binding of the first oligonucleotide to the second oligonucleotide represses expression of the target gene, and wherein nucleotide sequence of the second nucleotide is selected from the group consisting of SEQ ID NOs: 5, 8, 11, 14, 18, 24, 26, 27, 28, 29, 30, 33, 35, 36 and 4028-4238.

[0091] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of HEXA gene, wherein binding of the oligonucleotide to the mRNA transcript represses expression of HEXA gene.

[0092] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which is endogenously processed from a hairpin-shaped precursor, and anneals to a portion of a mRNA transcript of HEXA gene, wherein binding of the oligonucleotide to the mRNA transcript represses expression of HEXA gene.

[0093] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of HEXA gene, wherein binding of the oligonucleotide to the mRNA transcript represses expression of HEXA gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of: (a) a sequence selected from the group consisting of SEQ ID NOs: 4,8,23 and 4701-4737, and (b) the complement of a sequence selected from the group consisting of SEQ ID NOs: 8400, 8401 and 8382-8405.

[0094] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable oligonucleotide having a nucleotide sequence which has at least 80% sequence identity with a

nucleotide sequence selected from the group consisting of: (a) a sequence selected from the group consisting of SEQ ID NOs: 4,8,23 and 4701-4737, and (b) the complement of a sequence selected from the group consisting of SEQ ID NOs: 8400, 8401 and 8382-8405.

[0095] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of HEXA gene, wherein binding of the oligonucleotide to the mRNA transcript represses expression of HEXA gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs: 4,8,23 and 4701-4737.

[0096] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which is endogenously processed from a hairpin-shaped precursor, and anneals to a complementary portion of a mRNA transcript of HEXA gene, wherein binding of the oligonucleotide to the mRNA transcript represses expression of HEXA gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence se-

lected from the group consisting of SEQ ID NOs: 4,8,23 and 4701-4737, and.

[0097] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which includes from about 19 to about 22 nucleotides, which nucleotides anneal to a portion of a mRNA transcript of HEXA gene, and wherein the oligonucleotide is endogenously processed from a hairpin-form precursor, and includes at least 19 contiguous nucleotides from a sequence selected from the group consisting of SEQ ID NOs: 4,8,23 and 4701-4737.

[0098] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable first oligonucleotide which is a portion of a mRNA transcript of HEXA gene, and anneals to a second oligonucleotide that is endogenously processed from a hairpin precursor, wherein binding of the first oligonucleotide to the second oligonucleotide represses expression of the target gene.

[0099] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable first oligonucleotide which is a por-

tion of a mRNA transcript of HEXA gene, and anneals to a second oligonucleotide that is endogenously processed from a hairpin precursor, wherein binding of the first oligonucleotide to the second oligonucleotide represses expression of the target gene, and wherein nucleotide sequence of the second nucleotide is selected from the group consisting of SEQ ID NOs: 4,8,23 and 4701-4737.

[0100] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of APP gene, wherein binding of the oligonucleotide to the mRNA transcript represses expression of APP gene.

[0101] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which is endogenously processed from a hairpin-shaped precursor, and anneals to a portion of a mRNA transcript of APP gene, wherein binding of the oligonucleotide to the mRNA transcript represses expression of APP gene.

[0102] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which an-

neals to a portion of a mRNA transcript of APP gene, wherein binding of the oligonucleotide to the mRNA transcript represses expression of APP gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of: (a) a sequence selected from the group consisting of SEQ ID NOs: 2, 34 and 1757-1768, and (b) the complement of a sequence selected from the group consisting of SEQ ID NOs: 5097, 5104, 32101, 120890 and 5097-5112.

[0103] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of APP gene, wherein binding of the oligonucleotide to the mRNA transcript represses expression of APP gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs: 2, 34 and 1757-1768.

[0104] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which is endogenously processed from a hairpin-shaped precursor, and anneals to a complementary portion of a mRNA tran-

script of APP gene, wherein binding of the oligonucleotide to the mRNA transcript represses expression of APP gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs: 2, 34 and 1757-1768.

[0105] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which includes from about 19 to about 22 nucleotides, which nucleotides anneal to a portion of a mRNA transcript of APP gene, and wherein the oligonucleotide is endogenously processed from a hairpin-form precursor, and includes at least 19 contiguous nucleotides from a sequence selected from the group consisting of SEQ ID NOs: 2, 34 and 1757-1768.

[0106] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable first oligonucleotide which is a portion of a mRNA transcript of APP gene, and anneals to a second oligonucleotide that is endogenously processed from a hairpin precursor, wherein binding of the first oligonucleotide to the second oligonucleotide represses

expression of the target gene.

[0107] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable first oligonucleotide which is a portion of a mRNA transcript of APP gene, and anneals to a second oligonucleotide that is endogenously processed from a hairpin precursor, wherein binding of the first oligonucleotide to the second oligonucleotide represses expression of the target gene, and wherein nucleotide sequence of the second nucleotide is selected from the group consisting of SEQ ID NOs: 2, 34 and 1757-1768.

[0108] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated polynucleotide which is endogenously processed into a plurality of hairpin shaped precursor oligonucleotides, each of which is endogenously processed into a respective oligonucleotide, which in turn anneals to a portion of a mRNA transcript of a target gene, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene.

[0109] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which is en-

dogenously processed from a hairpin-shaped precursor, and anneals to a portion of a mRNA transcript of a target gene, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the target gene does not encode a protein.

[0110] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which is endogenously processed from a hairpin-shaped precursor, and anneals to a portion of a mRNA transcript of a target gene, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein a function of the oligonucleotide includes modulation of cell type.

[0111] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which is endogenously processed from a hairpin-shaped precursor, and anneals to a portion of a mRNA transcript of a target gene, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide is maternally transferred by a cell to at least one daughter cell of the cell, and a function

of the oligonucleotide includes modulation of cell type of the daughter cell.

[0112] There is moreover provided in accordance with another preferred embodiment of the present invention a method for bioinformatic detection of microRNA oligonucleotides, the method including: bioinformatically detecting a hairpin shaped precursor oligonucleotide, bioinformatically detecting an oligonucleotide which is endogenously processed from the hairpin shaped precursor oligonucleotide, and bioinformatically detecting a target gene of the oligonucleotide wherein the oligonucleotide anneals to at least one portion of a mRNA transcript of the target gene, and wherein the binding represses expression of the target gene, and the target gene is associated with a disease.

Brief Description of Drawings

[0113] Fig. 1 is a simplified diagram illustrating a genomic differentiation enigma that the present invention addresses;

[0114] Figs. 2, 3 and 4 are schematic diagrams which, when taken together, provide an analogy that illustrates a conceptual model of the present invention, addressing the genomic differentiation enigma;

[0115] Figs. 5A and 5B are schematic diagrams, which when taken together, illustrate a `genomic records` concept of

the conceptual model of the present invention, addressing the genomic differentiation enigma;

[0116] Fig. 6 is a schematic diagram illustrating a 'genomically programmed cell differentiation` concept of the conceptual model of the present invention, addressing the genomic differentiation enigma;

[0117] Fig. 7 is a schematic diagram illustrating a `genomically programmed cell-specific protein expression modulation` concept of the conceptual model of the present invention, addressing the genomic differentiation enigma;

[0118] Fig. 8 is a simplified diagram illustrating a mode by which an oligonucleotide of a novel group of oligonucleotides of the present invention, modulates expression of known target genes;

[0119] Fig. 9 is a simplified block diagram illustrating a bioinformatic oligonucleotide detection system capable of detecting oligonucleotides of the novel group of oligonucleotides of the present invention, which system is constructed and operative in accordance with a preferred embodiment of the present invention;

[0120] Fig. 10 is a simplified flowchart illustrating operation of a mechanism for training of a computer system to recognize the novel oligonucleotides of the present invention,

which mechanism is constructed and operative in accordance with a preferred embodiment of the present invention;

[0121] Fig. 11A is a simplified block diagram of a non-coding genomic sequence detector constructed and operative in accordance with a preferred embodiment of the present invention;

[0122] Fig. 11B is a simplified flowchart illustrating operation of a non-coding genomic sequence detector constructed and operative in accordance with a preferred embodiment of the present invention;

[0123] Fig. 12A is a simplified block diagram of a hairpin detector constructed and operative in accordance with a preferred embodiment of the present invention;

[0124] Fig. 12B is a simplified flowchart illustrating operation of a hairpin detector constructed and operative in accordance with a preferred embodiment of the present invention;

[0125] Fig. 13A is a simplified block diagram of a dicer-cut location detector constructed and operative in accordance with a preferred embodiment of the present invention;

[0126] Fig. 13B is a simplified flowchart illustrating training of a dicer-cut location detector constructed and operative in accordance with a preferred embodiment of the present

invention;

[0127] Fig. 13C is a simplified flowchart illustrating operation of a dicer-cut location detector constructed and operative in accordance with a preferred embodiment of the present invention;

[0128] Fig. 14A is a simplified block diagram of a target gene binding-site detector constructed and operative in accordance with a preferred embodiment of the present invention;

[0129] Fig. 14B is a simplified flowchart illustrating operation of a target gene binding-site detector constructed and operative in accordance with a preferred embodiment of the present invention;

[0130] Fig. 15 is a simplified flowchart illustrating operation of a function & utility analyzer constructed and operative in accordance with a preferred embodiment of the present invention;

[0131] Fig. 16 is a simplified diagram describing a novel bioinformatically detected group of regulatory polynucleotides referred to here as Genomic Record (GR) polynucleotide, each of which encodes an `operon-like` cluster of novel miRNA-like oligonucleotides, which in turn modulates expression of one or more target genes;

[0132] Fig. 17 is a simplified diagram illustrating a mode by which oligonucleotides of a novel group of operon-like polynucleotide of the present invention, modulate expression of other such polyonucleotides, in a cascading manner;

[0133] Fig. 18 is a block diagram illustrating an overview of a methodology for finding novel oligonucleotides and novel operon-like polynucleotides of the present invention, and their respective functions;

[0134] Fig. 19 is a block diagram illustrating different utilities of novel oligonucleotides and novel operon-like polynucleotides, both of the present invention;

[0135] Figs. 20A and 20B are simplified diagrams, which when taken together illustrate a mode of oligonucleotide-therapy applicable to novel oligonucleotides of the present invention;

[0136] Fig.21A is a histogram representing the distribution of known miRNA oligonucleotides, and that of miRNA-like hairpin-shaped oligonucleotides, predicted by the bioinformatics detection engine of the present invention, extracted from expressed genome sequences with respect to their hairpin detector score.

[0137] Fig. 21B is a table summarizing laboratory validation re-

sults which validate efficacy of a bioinformatic oligonucleotide detection system constructed and operative in accordance with a preferred embodiment of the present invention;

[0138] Fig. 22A and Fig.22B are a picture and a summary table of laboratory results validating the expression of 43 novel oligonucleotides detected by a bioinformatic oligonucleotide detection engine constructed and operative in accordance with a preferred embodiment of the present invention, thereby validating the efficacy of the oligonucleotide detection engine of the present invention;

[0139] Fig. 23A is a schematic representation of an "operon-like" cluster of novel hairpin sequences detected bioinformatically by a bioinformatic oligonucleotide detection engine constructed and operative in accordance with a preferred embodiment of the present invention, and non-GAM hairpin useful as negative controls thereto;

[0140] Fig. 23B is a schematic representation of secondary folding of hairpins of the operon-like cluster of Fig. 23A;

[0141] Fig. 23C is a picture of laboratory results demonstrating expression of novel oligonucleotides of Figs. 23A and 23B, and lack of expression of the negative controls, thereby validating efficacy of bioinformatic detection of

GAM oligonucleotides and GR polynucleotides of the present invention, by a bioinformatic oligonucleotide detection engine constructed and operative in accordance with a preferred embodiment of the present invention;

[0142] Fig. 24A, is an annotated sequence of EST72223 comprising known miRNA oligonucleotide MIR98 and novel oligonucleotide GAM25 PRECURSOR detected by the oligonucleotide detection system of the present invention; and

[0143] Figs. 24B, 24C and 24D are pictures of laboratory results, which when taken together demonstrate laboratory confirmation of expression of known oligonucleotide MIR98 and of novel bioinformatically detected GAM25 RNA respectively, both of Fig. 24A, thus validating the bioinformatic oligonucleotide detection system of the present invention.

Brief Description of Sequences

[0144] A Sequence Listing of genomic sequences of the present invention designated SEQ ID NO: 1 through SEQ ID NO: 142,621 is attached to this application, and is hereby incorporated herein. The genomic listing comprises the following nucleotide sequences: nucleotide sequences of 2147 GAMs precursors of respective novel oligonu-

cleotides of the present invention; nucleotide sequences of 4737 GAM RNA oligonucleotides of respective novel DNA oligonucleotides of the present invention; and nucleotide sequences of 135,737 target gene binding sites of respective novel oligonucleotides of the present invention.

Detailed Description

[0145] Reference is now made to Fig. 1 which is a simplified diagram providing a conceptual explanation of a genomic differentiation enigma, which the present invention addresses, inter alia.

[0146] Fig. 1 depicts various types of cells in an organism, such as a cartilage cell designated by reference numeral 1, a liver cell designated by reference numeral 2, a fibroblast cell designated by reference numeral 3, and a bone cell designated by reference numeral 4, all containing identical DNA designated by reference numeral 5. Notwithstanding that the various types of cells are all derived from an initial fertilized egg cell designated by reference numeral 6, each of these cells expresses different proteins and accordingly acquires a different shape and function.

[0147] The present invention proposes inter alia that the inevitable conclusion from the foregoing is, however, strik-

ingly simple: The genome must contain a modular differentiation coding system. The genome of each cell must include multiple modules or records, possibly a different one for each cell type, as well as a mechanism causing each cell at its inception to be instructed which one of the multiple records governs its behavior.

[0148] This modular code concept may be somewhat difficult to grasp, since most persons are accustomed to view things from an external viewpoint. An architect, for example, looks at a plan of a building, which details exactly where each element (block, window, door, electrical switch, etc.) is to be placed relative to all other elements, and, using the plan, instructs builders to place these elements in their designated places. This is an example of an external viewpoint: The architect is external to the plan, which itself is external with respect to the physical building, and with respect to its various elements. The architect may therefore act as an "external organizing agent": seeing the full picture and the relationships between all elements, and being able to instruct from the outside where to place each of them.

[0149] According to a preferred embodiment of the present invention, genomic differentiation coding works differently,

without any such external organizing agent. It comprises a smart block (the first cell), which is the architect and the plan, and which continuously duplicates itself, somehow knowing when to manifest itself as a block and when as a window, door, or electrical switch.

[0150] Reference is now made to Figs. 2A – 4 which are schematic diagrams which, when taken together, provide an analogy that illustrates a conceptual model of the present invention, addressing the genomic differentiation enigma.

[0151] Reference is now made to Fig. 2A. An imaginary talented chef, designated by reference numeral 7, is capable of preparing any meal provided that he is given specific written cooking instructions. This chef 7 is equipped with two items: (a) a thick recipe book, designated by reference numeral 8, and (b) a small note, designated by reference numeral 9, having a number scribbled on it. The recipe book 8 comprises multiple pages, each page detailing how to prepare a specific meal. The small note 9 indicates the page to be opened, and therefore the meal to be prepared. The chef looks at the page number written on the note, opens the recipe book to the appropriate page, and prepares the meal according to the written instructions on

this page. In the example shown in Fig. 2A, the chef 7 is holding a small note 9 bearing the number 127. He therefore opens the book to page 127, as designated by reference numeral 10. Since this page contains the recipe for preparing bread, the chef 7 prepares a loaf of bread, designated by reference numeral 12. Pages of the book, such as page 10 in the example shown in Fig. 2A, contains additional information, designated by reference numeral 11 which additional data is further elaborated hereinbelow with reference to Figs. 3 and 4.

[0152] Reference is now made to Fig. 2B, which depicts two identical chefs, a first chef, designated by reference numeral 13, and a second chef, designated by reference numeral 14, both holding an identical recipe book, designated by reference numeral 8. Although the first chef 13 and the second chef 14 are identical, and hold identical recipe books 8, they differ in that they hold different small notes: the first chef 13 holds a small note designated by reference numeral 9, having the number 127 written on it, whereas the second chef 14 holds a small note designated by reference numeral 15, having the number 134 written on it. Accordingly, the first chef 13 opens the book 8 to page 127, as designated by reference numeral 10 and,

based on the instructions written on page 127 prepares a loaf of bread, designated by reference numeral 12. The second chef 14 opens the book 8 to page 134, as designated by reference numeral 16 and, based on the instructions written on page 134, prepares a pie, designated by reference numeral 17. Pages in the book, such as pages 10 and 16 in the examples shown in Fig. 2B, contain additional information, designated by reference numeral 11 which additional information is further elaborated hereinbelow with reference to Figs. 3 and 4.

[0153] Reference is now made to Fig. 3 which illustrates a mode by which an imaginary chef can duplicate himself yielding two identical chefs, instructing each of the identical duplicate chefs to prepare a different meal. As an example, Fig. 3 shows chef 21 duplicating himself, yielding two duplicate chefs: a first duplicate chef designated by reference numeral 22 and a second duplicate chef designated by reference numeral 23. The duplicate chefs are identical to each other and to chef 21.

[0154] Like chefs 7 and 13 (Fig. 2A and 2B), Fig. 3 shows chef 21 holding a recipe book 8 and receiving a note 9 bearing the number 127. The chef 21 therefore opens the book 8 to page 127, as designated by reference numeral 10, and

prepares a loaf of bread 12. However, Fig. 3 also elaborates some of the additional information 11 (Figs. 2A and 2B) found in page 10: the bottom of page 10, bears two numbers, 134 and 157.

[0155] Chef 21 is trained to perform the following three actions when he is finished preparing a meal: (a) Duplicate himself yielding two duplicate chefs, the first duplicate chef 22 and the second duplicate chef 23; (b) Duplicate his recipe book 8, handing an identical copy to each of the duplicate chefs 22 and 23; and (c) Write down the numbers found at the bottom of the page he was instructed to open the book to. In the example of chef 21, since he was instructed to open the book to page 10, he writes the numbers 134 and 157 on two respective notes designated by reference numerals 15 and 24, and hands note 15 bearing the number 134 to the first duplicate chef 22 and note 24 bearing the number 157 to the second duplicate chef 23.

[0156] Accordingly, the first duplicate chef 22 receives note 15 bearing the number 134 and therefore opens the recipe book 8 to page 134, as designated by reference numeral 16, and prepares a pie, designated by reference numeral 17. The second duplicate chef 23 receives note 24 bearing the number 157 and therefore opens the recipe book 8 to

page 157, as designated by reference numeral 25, and prepares rice, designated by reference numeral 26.

[0157] It is appreciated that while chef 21 and duplicate chefs 22 and 23 are identical and hold identical recipe books 8, they each prepare a different meal. It is also appreciated that the meals prepared by the first duplicate chef 22 and the second duplicate chef 23 are determined by chef 21, and are mediated by the differently numbered notes 15 and 24 passed on from chef 21 to duplicate chefs 22 and 23 respectively.

[0158] It is further appreciated that the mechanism illustrated by Fig. 3 enables an unlimited lineage of chefs to divide into duplicate, identical chefs and to determine the meals those duplicate chefs would prepare. As an example, since the first duplicate chef 22 is directed to page 134, as designated by reference numeral 16, when he duplicates himself (not shown), he will instruct his two duplicate chefs to prepare meals specified on pages the numbers of which are written at the bottom of page 134, i.e. pages 114 and 193 respectively. Similarly, the second duplicate chef 23 will instruct its duplicate chefs to prepare meals specified on pages 121 and 146 respectively, etc.

[0159] Reference is now made to Fig. 4, which illustrates a mode

by which a chef can prepare a meal based on instructions written in a shorthand format: The main meal-page to which a chef is directed by a small note he is given, merely contains a list of numbers which further direct him to other pages, each specifying how to prepare an ingredient of that meal.

[0160] To illustrate this shorthand format Fig. 4 shows a chef, designated by reference numeral 27, holding the recipe book 8 and the note 9 which bears the number 127. The chef 27 accordingly opens the recipe book 8 to page 127, as designated by reference numeral 10, and based on instructions on this page prepares bread 12. This is similar to chefs 7, 13 and 21 of Figs. 2A, 2B and 3 respectively.

[0161] However, Fig. 4 also further elaborates some of the additional information 11 (Figs. 2A and 2B) found in page 10. Fig. 4 shows the cooking "instructions" found on page 10 for making bread 12 written in a shorthand format, comprising only three numbers, 118, 175 and 183. The chef 27 writes these numbers on three respective notes designated by reference numerals 28 - 30. The notes 28 -30 are then used to turn to corresponding pages 31 - 33 of the book 8, which pages provide instructions for preparation of ingredients required for making bread 12: flour 34,

milk 35 and salt 36.

[0162] The analogy provided by Figs. 2A – 4 illustrates the conceptual model of the present invention addressing the genomic differentiation enigma, and may be explained as follows: The chefs and duplicate chefs 7, 13, 14, 21 – 23 and 27 (Figs. 2A – 4) in the given analogy represents cells. The thick recipe book 8 represents the DNA 5 (Fig. 1). Preparing meals such as bread 12, pie 17 or rice 16 (all of Fig. 3) represent the cell manifesting itself as a specific cell-type, such as cartilage cell 1, liver cell 2, fibroblast cell 3, and bone cell 4 (all of Fig. 1). Ingredients of a meal, such as flour 34, milk 35 and salt 36 (all of Fig. 4), represent proteins typically expressed by a cell of a certain cell-type, such as 1 – 4. Like the different chefs of the analogy, having the same thick recipe book 8 yet preparing different meals, so do different cells in an organism contain the same DNA 5 yet manifest themselves as different cell types, such as 1 – 4, expressing proteins typical of these respective cell types. Application of analogy of Figs. 2A – 4 to cell-biology is further described hereinbelow with reference to Figs. 5A – 7.

[0163] Reference is now made to Figs. 5A and 5B which are schematic diagrams, which when taken together illustrate

a Genomic Records concept of the present invention, addressing the genomic differentiation enigma. Figs. 5A and 5B correspond to Figs. 2A and 2B of the chef analogy described hereinabove.

[0164] An important aspect of the present invention is the Genomic Records concept. According to a preferred embodiment of the present invention the DNA (the thick recipe book 8 in the illustration) comprises a very large number of Genomic Records (analogous to pages, such as 10, 16 and 25, in the recipe book) containing the instructions for differentiation of a different cell-type, or developmental process. Each Genomic Record comprises by a very short genomic sequence which functions as a "Genomic Address" of that Genomic Record (analogous to a page number, such as the numbers 127, 134 and 157 appearing in Fig. 3, in the recipe book). At its inception, in addition to the DNA, each cell also receives a short RNA segment (the scribbled short note, such as 9, 15, 24 of Fig. 3 in the illustration). This short RNA segment binds complementarily to a "Genomic Address" sequence of one of the Genomic Records, thereby modulating expression of that Genomic Record, and accordingly determining the cell's-fate (analogous to opening the recipe book 8 to a

page corresponding to a number on the scribbled note, thereby determining the meal to be prepared). A Genomic Record may also comprise multiple short RNA segments each of which binds complementarily to a target protein coding gene, thus modulating expression of this target gene (analogous to the shorthand format whereby a page, such as 10, points to other pages, such as 31 -33, encoding various ingredient, such as 34, 35 and 36, all of Fig. 4).

[0165] Reference is now made to Fig. 5A. Fig 5A illustrates a cell 37, having a genome 38. The genome 38 comprises a plurality of Genomic Records, some of which Genomic Records correlate to specific cell-types. As an example, 6 such genomic records are shown, corresponding to 6 cell-types: LYMPH genomic record 39, FIBROBLAST genomic record 40, MUSCLE genomic record 41, BONE genomic record 42, CARTILAGE genomic record 43 and NERVE genomic record 44. Each genomic record comprises genomic instructions on differentiation into a specific cell-type, as further elaborated hereinbelow with reference to Fig. 7. At cell inception, the cell 37 receives a maternal short RNA segment 46, which activates one of the genomic records, causing the cell to differentiate according to the instruc-

tions this genomic record comprises. As an example Fig. 5A illustrates reception of a maternal short RNA segment designated 46 having a nucleotide sequence herein symbolically represented by A' .

[0166] The FIBROBLAST genomic record 40 contains a binding site having a nucleotide sequence symbolically represented by A, which is complementary to the nucleotide sequence of A' , and therefore the short RNA segment 46 binds to the FIBROBLAST genomic record 40. This binding activates the FIBROBLAST genomic record, causing the cell 37 to differentiate into a fibroblast cell-type 3 (Fig. 1). Other genomic records, designated by reference numerals 39 and 41 - 44, comprise binding sites having nucleotide sequences that are symbolically represented by F, E, B, C and D, which are not complementary of the nucleotide sequence of the short RNA segment 46, and are therefore not activated thereby. Genomic Records, such as the FIBROBLAST genomic record 40 contain additional information, designated by reference numeral 45, which is further elaborated hereinbelow with reference to Figs. 6 and 7.

[0167] Reference is now made to Fig. 5B, which is a simplified schematic diagram, illustrating cellular differentiation mediated by the "Genomic Records" concept. Fig. 5B depicts

2 cells in an organism, CELL A designated by reference numeral 47 and CELL B designated by reference numeral 48, each having a genome 38. It is appreciated that since CELL A 47 and CELL B 48 are cells in the same organism, the genome 38 of cells 47 and 48 is identical. Despite having an identical genome 38, CELL A 47 differentiates differently from CELL B 48, due to activation of different genomic records in these two cells. In CELL A 47 the FIBRO GENOMIC RECORD 40 is activated, causing CELL A 47 to differentiate into a FIBROBLAST CELL 3, whereas in CELL B 48 the BONE GENOMIC RECORD 42 is activated, causing the CELL B 48 to differentiate into a BONE CELL 4 (Fig. 1). The cause for activation of different genomic records in these two cells is the different maternal short RNA which they both received: CELL A 47 received a maternal short RNA segment designated 46 bearing a nucleotide sequence represented by A' activating genomic record FIBRO 40, whereas CELL B 48 received a maternal short RNA segment designated 49 bearing a nucleotide sequence represented by B' activating genomic record BONE 42.

[0168] Reference is now made to Fig. 6 which is a schematic diagram illustrating a "genomically programmed cell differentiation" concept of the conceptual model of the present

invention, addressing the genomic differentiation enigma.

[0169] A cell designated CELL A 50 divides into 2 cells designated CELL B 51 and CELL C 52. CELL A 50, CELL B 51 and CELL C 52 each comprise a GENOME 38, which GENOME 38 comprises a plurality of GENOMIC RECORDS, herein exemplified by reference numerals 40, 42 and 43. It is appreciated that since CELL A 50, CELL B 51 and CELL C 52 are cells in the same organism, the GENOME 38 of these cells, and the GENOMIC RECORDS, exemplified by 40, 42 and 43, the genome of these cells comprises, are identical in these cells.

[0170] As described above with reference to Fig. 5B, at its inception, CELL A 50 receives a maternal short RNA segment, designated by reference numeral 46, having a nucleotide sequence represented by A' and outlined by a broken line, which activates the FIBRO genomic record 40, thereby causing CELL A 50 to differentiate into a FIBROBLAST CELL 3. However, Fig. 6 elaborates some of the additional information 45 (Fig. 5A) of the genomic records: Genomic record may also comprise two short genomic sequences, referred to here as Daughter Cell Genomic Addresses. Blocks designated B and C are Daughter Cell Genomic Addresses of the FIBRO Genomic Record. At cell division,

each parent cell transcribes two short RNA segments, corresponding to the two Daughter Cell Genomic Addresses of the Genomic Record of that parent cell, and transfers one to each of its two daughter cells. CELL A 50 transcribes and transfers to its two daughter cells 51 and 52, two short RNA segments, designated by reference numerals 49 and 53, outlined by a broken line and designated B' and C' , corresponding to daughter cell genomic addresses designated B and C comprised in the FIBRO genomic record 40.

[0171] CELL B 51 therefore receives the above mentioned maternal short RNA segment designated 49 having a nucleotide sequence represented by B' , which binds complementarily to genomic address designated B of the BONE genomic record 42, thereby activating this genomic record, which in turn causes CELL B 51 to differentiate into a BONE CELL 4. Similarly, CELL C 52 receives the above mentioned maternal short RNA segment designated 53 having a nucleotide sequence represented by C' , which binds complementarily to genomic address designated C of a CARTILAGE genomic record 43, thereby activating this genomic record, which in turn causes CELL C 52 to differentiate into a CARTILAGE CELL 1 (Fig. 1).

[0172] It is appreciated that the mechanism illustrated by Fig. 6 enables an unlimited lineage of cells to divide into daughter cells containing the same DNA 5 (Fig. 1), and to determine the cell-fate of these daughter cells. For example, when CELL B 51 and CELL C 52 divide into their respective daughter cells (not shown), they will transfer short RNA segments designated by reference numerals 54 - 57, to their respective daughter cells. The cell fate of each of these daughter cells is effected by the identity of the maternal short RNA segments 54 - 57 they each receive, which in turn determine the genomic record activated.

[0173] Reference is now made to Fig. 7 which is a schematic diagram illustrating a "genomically programmed cell-specific protein expression modulation" concept of the conceptual model of the present invention, addressing the genomic differentiation enigma.

[0174] Cell A 58 receives a maternal short RNA segment designated 46 having a nucleotide sequence represented by A' , which activates the FIBROBLAST genomic record 40, by complementarily binding to a binding site this genomic record comprises, the nucleotide sequence of which binding site is designated A. This is similar to the process shown in Fig. 5A. However, Fig. 7 further elaborates some

of the additional information 45 (Fig. 1). The FIBROBLAST genomic record 40 comprises 3 short nucleotide segments, having nucleotide sequences symbolically represented by 1, 2 and 4 respectively, which encode 3 respective short RNA oligonucleotides, designated by reference numerals 59 - 61. Each of these short RNA oligonucleotides modulates expression of a respective one of the target genes GENE 1, GENE 2 and GENE 4, designated by reference numerals 62 - 64 respectively, by complementarily binding to a binding site sequence associated with that target gene. In a preferred embodiment of the present invention, the modulation of expression of target genes such as 62 - 64 comprises translation inhibition of target genes by complementarily binding to binding sites located in untranslated regions of the target genes. Modulation of expression of these genes results in CELL A 58 differentiating into a FIBROBLAST cell-type 3 (Fig. 1).

[0175] It is appreciated that the concept of genomic records each comprising a cluster of short RNA segments, which segments modulate expression of target genes thereby modulating differentiation, is compatible with the clusters of miRNA oligonucleotides of the present invention, and their translational inhibition of respective target genes by

means of complementarily binding to binding sites located in the untranslated regions of mRNA of these target genes.

[0176] Reference is now made to Fig. 8, which is a simplified diagram describing how a plurality of novel bioinformatically detectable oligonucleotides of the present invention, referred to here as Genomic Address Messenger (GAM) oligonucleotides, modulate expression of respective target genes.

[0177] GAM oligonucleotides are novel, bioinformatically detectable, regulatory, non protein coding, micro RNA (miRNA)-like oligonucleotides. The method by which GAMs are detected is described hereinbelow with additional reference to Figs. 9-15.

[0178] GAM PRECURSOR DNA is encoded by the human genome. GAM TARGET GENE is a human gene encoded by the human genome.

[0179] GAM PRECURSOR DNA encodes a GAM PRECURSOR RNA. Similar to miRNA oligonucleotides, GAM PRECURSOR RNA does not encode a protein. GAM PRECURSOR RNA folds onto itself, forming GAM FOLDED PRECURSOR RNA, which has a two-dimensional `hairpin structure`. As is well known in the art, this `hairpin structure`, is typical of by

miRNA precursor oligonucleotides, and is due to the fact that the nucleotide sequence of the first half of the miRNA precursor oligonucleotide is a fully or partially complementary sequence of the nucleotide sequence of the second half thereof. By complementary is meant a sequence which is reversed and wherein each nucleotide is replaced by a complementary nucleotide, as is well known in the art (e.g. ATGGC is the complementary sequence of GCCAT).

[0180] An enzyme complex comprising an enzyme called Dicer together with other necessary proteins, herein designated as the DICER COMPLEX, `dices` the GAM FOLDED PRECURSOR RNA yielding a GAM RNA, in the form of a single stranded ~22 nt long RNA segment. The DICER COMPLEX is known in the art to dice a hairpin structured miRNA precursor, thereby yielding diced miRNA in the form of a short ~22nt RNA segment.

[0181] GAM TARGET GENE encodes a corresponding messenger RNA, designated GAM TARGET RNA. GAM TARGET RNA comprises three regions, as is typical of mRNA of a protein coding gene: a 5` untranslated region, a protein coding region and a 3` untranslated region, designated 5`UTR, PROTEIN CODING and 3`UTR respectively.

[0182] GAM RNA binds complementarily (i.e. hybridizes) to one

or more target binding sites located in untranslated regions of GAM TARGET RNA. This complementary binding is due to the fact that the nucleotide sequence of GAM RNA is a partial or fully complementary sequence of the nucleotide sequence of each of the target binding sites. As an illustration, Fig. 8 shows three such target binding sites, designated BINDING SITE I, BINDING SITE II and BINDING SITE III respectively. It is appreciated that the number of target binding sites shown in Fig. 8 is only illustrative and that any suitable number of target binding sites may be present. It is further appreciated that although Fig. 8 shows target binding sites only in the 3`UTR region, these target binding sites may be located instead in the 5`UTR region or in both 3`UTR and 5`UTR regions.

[0183] The complementary binding of GAM RNA to target binding sites on GAM TARGET RNA, such as BINDING SITE I, BINDING SITE II and BINDING SITE III, inhibits translation of GAM TARGET RNA into GAM TARGET PROTEIN, which is shown surrounded by a broken line.

[0184] It is appreciated that GAM TARGET GENE in fact represents a plurality of GAM target genes. The mRNA of each one of this plurality of GAM target genes comprises one or more

target binding sites, each having a nucleotide sequence which is at least partly complementary to GAM RNA, and which when bound by GAM RNA causes inhibition of translation of the GAM target mRNA into a corresponding GAM target protein.

[0185] The mechanism of the translational inhibition exerted by GAM RNA on one or more GAM TARGET GENE, may be similar or identical to the known mechanism of translational inhibition exerted by known miRNA oligonucleotides.

[0186] Nucleotide sequences of each of a plurality of GAM oligonucleotides described by Fig. 8 and their respective genomic sources and chromosomal locations are set forth in Tables 1 – 3, hereby incorporated herein.

[0187] Nucleotide sequences of GAM PRECURSOR RNAs, and a schematic representation of a predicted secondary folding of GAM FOLDED PRECURSOR RNAs, of each of a plurality of GAM oligonucleotides described by Fig. 8 are set forth in Table 4, hereby incorporated herein.

[0188] Nucleotide sequences of a `diced` GAM RNA of each of a plurality of GAM oligonucleotides described by Fig. 8 are set forth in Table 5, hereby incorporated herein.

[0189] Nucleotide sequences of target binding sites, such as

BINDING SITE I, BINDING SITE II and BINDING SITE III found on GAM TARGET RNA, of each of a plurality of GAM oligonucleotides described by Fig. 8, and a schematic representation of the complementarity of each of these target binding sites to each of a plurality of GAM RNAs described by Fig. 8 are set forth in Tables 6 and 7, hereby incorporated herein.

[0190] It is appreciated that specific functions and accordingly utilities of each of a plurality of GAM oligonucleotides described by Fig. 8 correlate with, and may be deduced from, the identity of the GAM TARGET GENEs inhibited thereby, whose . functions are set forth in Table 8, hereby incorporated herein.

[0191] Studies documenting well known correlations between each of a plurality of GAM TARGET GENEs of the GAM oligonucleotides of Fig. 8, and known functions and diseases are listed in Table 9, hereby incorporated herein.

[0192] The present invention discloses a novel group of oligonucleotides, belonging to the miRNA-like oligonucleotides group, here termed GAM oligonucleotides, for which a specific complementary binding has been determined bioinformatically.

[0193] Reference is now made to Fig. 9 which is a simplified

block diagram illustrating a bioinformatic oligonucleotide detection system and method constructed and operative in accordance with a preferred embodiment of the present invention.

[0194] An important feature of the present invention is a BIOINFORMATIC OLIGONUCLEOTIDE DETECTION ENGINE 100, which is capable of bioinformatically detecting oligonucleotides of the present invention.

[0195] The functionality of the BIOINFORMATIC OLIGONUCLEOTIDE DETECTION ENGINE 100 includes receiving EXPRESSED RNA DATA 102, SEQUENCED DNA DATA 104, and PROTEIN FUNCTION DATA 106; performing a complex process of analysis of this data as elaborated hereinbelow, and based on this analysis provides information, designated by reference numeral 108, identifying and describing features of novel oligonucleotides.

[0196] EXPRESSED RNA DATA 102 comprises published expressed sequence tags (EST) data, published mRNA data, as well as other published RNA data. SEQUENCED DNA DATA 104 comprises alphanumeric data representing genomic sequences and preferably including annotations such as information indicating the location of known protein coding regions relative to the genomic sequences.

[0197] PROTEIN FUNCTION DATA 106 comprises information from scientific publications e.g. physiological functions of known proteins and their connection, involvement and possible utility in treatment and diagnosis of various diseases.

[0198] EXPRESSED RNA DATA 102 and SEQUENCED DNA DATA 104 may preferably be obtained from data published by the National Center for Biotechnology Informatiion (NCBI) at the National Institute of Health (NIH) (Jenuth,J.P. (2000). Methods Mol. Biol. 132:301-312(2000) , herein incorporated by reference).

[0199] , as well as from various other published data sources. PROTEIN FUNCTION DATA 106 may preferably be obtained from any one of numerous relevant published data sources, such as the Online Mendelian Inherited Disease In Man (OMIM(TM), Hamosh et al., Nucleic Acids Res. 30: 52-55(2002)) database developed by John Hopkins University, and also published by NCBI (2000).

[0200] Prior to or during actual detection of BIOINFORMATICALLY DETECTED GROUP OF NOVEL OLIGONUCLEOTIDES 108 by the BIOINFORMATIC OLIGONUCLEOTIDE DETECTION ENGINE 100, BIOINFORMATIC OLIGONUCLEOTIDE DETECTION ENGINE TRAINING & VALIDATION FUNCTIONALITY 110 is

operative. This functionality uses one or more known miRNA oligonucleotides as a training set to train the BIOINFORMATIC OLIGONUCLEOTIDE DETECTION ENGINE 100 to bioinformatically recognize miRNA-like oligonucleotides, and their respective potential target binding sites. BIOINFORMATIC OLIGONUCLEOTIDE DETECTION ENGINE TRAINING &VALIDATION FUNCTIONALITY 110 is further described hereinbelow with reference to Fig. 10.

[0201] The BIOINFORMATIC OLIGONUCLEOTIDE DETECTION ENGINE 100 preferably comprises several modules which are preferably activated sequentially, and are described as follows:

[0202] A NON-CODING GENOMIC SEQUENCE DETECTOR 112 operative to bioinformatically detect non-protein coding genomic sequences. The NON-CODING GENOMIC SEQUENCE DETECTOR 112 is further described herein below with reference to Figs. 11A and 11B.

[0203] A HAIRPIN DETECTOR 114 operative to bioinformatically detect genomic `hairpin-shaped` sequences, similar to GAM FOLDED PRECURSOR RNA (Fig.8). The HAIRPIN DETECTOR 114 is further described herein below with reference to Figs. 12A and 12B.

[0204] A DICER-CUT LOCATION DETECTOR 116 operative to

bioinformatically detect the location on a GAM FOLDED PRECURSOR RNA which is enzymatically cut by DICER COMPLEX (Fig.8), yielding diced GAM RNA. The DICER-CUT LOCATION DETECTOR 116 is further described herein below with reference to Figs. 13A - 13C.

[0205] A TARGET GENE BINDING-SITE DETECTOR 118 operative to bioinformatically detect target genes having binding sites, the nucleotide sequence of which is partially complementary to that of a given genomic sequence, such as a nucleotide sequence cut by DICER COMPLEX. The TARGET GENE BINDING-SITE DETECTOR 118 is further described hereinbelow with reference to Figs. 14A and 14B.

[0206] A FUNCTION & UTILITY ANALYZER 120 operative to analyze function and utility of target genes, in order to identify target genes which have a significant clinical function and utility. The FUNCTION & UTILITY ANALYZER 120 is further described hereinbelow with reference to Fig. 15.

[0207] According to a preferred embodiment of the present invention the engine 100 may employ a cluster of 40 PCs (XEON (R) , 2.8GHz, with 80GB storage each), connected by Ethernet to 8 servers (2-CPU, XEON (TM) 1.2-2.2GHz, with ~200GB storage each), combined with an 8-processor server (8-CPU, Xeon 550Mhz w/ 8GB RAM) connected via

2 HBA fiber-channels to an EMC CLARIION (TM) 100-disks, 3.6 Terabyte storage device. A preferred embodiment of the present invention may also preferably comprise software which utilizes a commercial database software program, such as MICROSOFT (TM) SQL Server 2000. It is appreciated that the above mentioned hardware configuration is not meant to be limiting, and is given as an illustration only. The present invention may be implemented in a wide variety of hardware and software configurations.

[0208] The present invention discloses 2147 novel oligonucleotides of the GAM group of oligonucleotides, which have been detected bioinformatically, as set forth in Tables 1 - 4, and 313 novel polynucleotides of the GR group of polynucleotides , which have been detected bioinformatically. Laboratory confirmation of 43 bioinformatically predicted oligonucleotides of the GAM group of oligonucleotides , and several bioinformatically predicted polynucleotides of the GR group of polynucleotides, is described hereinbelow with reference to Figs. 21 - 24D.

[0209] Reference is now made to Fig. 10 which is a simplified flowchart illustrating operation of a preferred embodiment of the BIOINFORMATIC OLIGONUCLEOTIDE DETECTION

ENGINE TRAINING &VALIDATION FUNCTIONALITY 110 described hereinabove with reference to Fig. 9.

[0210] BIOINFORMATIC OLIGONUCLEOTIDE DETECTION ENGINE TRAINING & VALIDATION FUNCTIONALITY 110 begins by training the BIOINFORMATIC OLIGONUCLEOTIDE DETECTION ENGINE 100 (Fig.9) to recognize one or more known miRNA oligonucleotides, as designated by reference numeral 122. This training step comprises HAIRPIN DETECTOR TRAINING & VALIDATION FUNCTIONALITY 124, further described hereinbelow with reference to Fig. 12A, DICER-CUT LOCATION DETECTOR TRAINING & VALIDATION FUNCTIONALITY 126, further described hereinbelow with reference to Fig. 13A and 13B, and TARGET GENE BINDING-SITE DETECTOR TRAINING & VALIDATION FUNCTIONALITY 128, further described hereinbelow with reference to Fig. 14A.

[0211] Next, the BIOINFORMATIC OLIGONUCLEOTIDE DETECTION ENGINE TRAINING & VALIDATION FUNCITONALITY 110 is operative bioinformatically detect novel oligonucleotides, using BIOINFORMATIC OLIGONUCLEOTIDE DETECTION ENGINE 100 (Fig.9), as designated by reference numeral 130. Wet lab experiments are preferably conducted in order to validate expression and preferably function of some sam-

ples of the novel oligonucleotides detected by the BIOINFORMATIC OLIGONUCLEOTIDE DETECTION ENGINE 100, as designated by reference numeral 132. Figs. 22A - 24D illustrate examples of wet-lab validation of the above mentioned sample novel oligonucleotides bioinformatically detected in accordance with a preferred embodiment of the present invention.

[0212] Reference is now made to Fig. 11A which is a simplified block diagram of a preferred implementation of the NON-CODING GENOMIC SEQUENCE DETECTOR 112 described hereinabove with reference to Fig. 9. The NON-PROTEIN CODING GENOMIC SEQUENCE DETECTOR 112 preferably receives at least two types of published genomic data: EXPRESSED RNA DATA 102 and SEQUENCED DNA DATA 104. The EXPRESSED RNA DATA 102 may include, inter alia, EST data, EST clusters data, EST genome alignment data and mRNA data. Sources for EXPRESSED RNA DATA 102 include NCBI dbEST, NCBI UniGene clusters and mapping data, and TIGR (Kirkness F. and Kerlavage,A.R., Methods Mol. Biol. 69:261-268 (1997))

[0213] gene indices. SEQUENCED DNA DATA 104 may include sequence data (FASTA format files), and feature annotations (GenBank file format) mainly from NCBI databases. Based

on the above mentioned input data, the NON-PROTEIN CODING GENOMIC SEQUENCE DETECTOR 112 produces a plurality of NON-PROTEIN CODING GENOMIC SEQUENCES 136. Preferred operation of the NON-PROTEIN CODING GENOMIC SEQUENCE DETECTOR 112 is described hereinbelow with reference to Fig. 11B

[0214] Reference is now made to Fig. 11B which is a simplified flowchart illustrating a preferred operation of the NON-CODING GENOMIC SEQUENCE DETECTOR 112 of Fig. 9. Detection of NON-PROTEIN CODING GENOMIC SEQUENCES 136, generally preferably progresses along one of the following two paths:

[0215] A first path for detecting NON-PROTEIN CODING GENOMIC SEQUENCES 136 (Fig. 11A) begins with receipt of a plurality of known RNA sequences, such as EST data. Each RNA sequence is first compared with known protein-coding DNA sequences, in order to select only those RNA sequences which are non-protein coding, i.e. intergenic or intronic sequences. This can preferably be performed by using one of many alignment algorithms known in the art, such as BLAST (Altschul et al., J. Mol. Biol. 215:403-410 (1990)). This sequence comparison preferably also provides localization of the RNA sequence on the DNA se-

quences.

[0216] Alternatively, selection of non-protein coding RNA sequences and their localization on the DNA sequences can be performed by using publicly available EST cluster data and genomic mapping databases, such as the UNIGENE database published by NCBI or the TIGR database. Such databases, map expressed RNA sequences to DNA sequences encoding them, find the correct orientation of EST sequences, and indicate mapping of ESTs to protein coding DNA regions, as is well known in the art. Public databases, such as TIGR, may also be used to map an EST to a cluster of ESTs, known in the art as Tentative Human Consensus and assumed to be expressed as one segment.. Publicly available genome annotation databases, such as NCBIs GenBank, may also be used to deduce expressed intronic sequences.

[0217] Optionally, an attempt may be made to "expand" the non-protein RNA sequences thus found, by searching for transcription start and end signals, respectively upstream and downstream of the location of the RNA on the DNA, as is well known in the art.

[0218] A second path for detecting NON-PROTEIN CODING GENOMIC SEQUENCES 136 (Fig. 11A) begins with receipt of

DNA sequences. The DNA sequences are parsed into non protein coding sequences, using published DNA annotation data, by extracting those DNA sequences which are between known protein coding sequences. Next, transcription start and end signals are sought. If such signals are found, and depending on their robustness, probable expressed non-protein coding genomic sequences are obtained. Such approach is especially useful for identifying novel GAM oligonucleotides which are found in proximity to other known miRNA oligonucleotides, or other wet-lab validated GAM oligonucleotides . Since, as described hereinbelow with reference to Fig. 16, GAM oligonucleotides are frequently found in clusters, sequences located near known miRNA oligonucleotides are more likely to contain novel GAM oligonucleotides. Optionally, sequence orthology, i.e. sequence conservation in an evolutionary related species, may be used to select genomic sequences having a relatively high probability of containing expressed novel GAM oligonucleotides.

[0219] Reference is now made to Fig. 12A which is a simplified block diagram of a preferred implementation of the HAIRPIN DETECTOR 114 described hereinabove with reference to Fig. 9.

[0220] The goal of the HAIRPIN DETECTOR 114 is to detect hairpin-shaped genomic sequences, similar to those of known miRNA oligonucleotides. A hairpin- shaped genomic sequence is a genomic sequence, having a first half which is at least partially complementary to a second half thereof, which causes the halves to folds onto themselves, thereby forming a hairpin structure, as mentioned hereinabove with reference to Fig. 8.

[0221] The HAIRPIN DETECTOR 114 (Fig. 9) receives a plurality of NON-PROTEIN CODING GENOMIC SEQUENCES 136 (Fig. 11A). Following operation of HAIRPIN DETECTOR TRAINING & VALIDATION FUNCTIONALITY 124 (Fig. 10), the HAIRPIN DETECTOR 114 is operative to detect and output hairpin-shaped sequences, which are found in the NON-PROTEIN CODING GENOMIC SEQUENCES 136. The hairpin-shaped sequences detected by the HAIRPIN DETECTOR 114 are designated HAIRPINS STRUCTURES ON GENOMIC SEQUENCES 138. A preferred mode of operation of the HAIRPIN DETECTOR 114 is described hereinbelow with reference to Fig. 12B.

[0222] HAIRPIN DETECTOR TRAINING & VALIDATION FUNCTIONALITY 124 includes an iterative process of applying the HAIRPIN DETECTOR 114 to known hairpin shaped miRNA

precursor sequences, calibrating the HAIRPIN DETECTOR 114 such that it identifies a training set of known hairpin-shaped miRNA precursor sequences, as well as other similarly hairpin-shaped sequences. In a preferred embodiment of the present invention, the HAIRPIN DETECTOR TRAINING & VALIDATION FUNCTIONALITY 124 trains the HAIRPIN DETECTOR 114 and validates each of the steps of operation thereof described hereinbelow with reference to Fig. 12B.

[0223] The HAIRPIN DETECTOR TRAINING & VALIDATION FUNCTIONALITY 124 preferably uses two sets of data: the aforesaid training set of known hairpin-shaped miRNA precursor sequences, such as hairpin-shaped miRNA precursor sequences of 440 miRNA oligonucleotides of H. sapiens, M. musculus, C. elegans, C. Brigssae and D. Melanogaster, annotated in the RFAM database (Griffiths-Jones 2003), and a large background set of about 350,000 hairpin-shaped sequences found in expressed non-protein coding genomic sequences. The background set is expected to comprise some valid, previously undetected hairpin-shaped miRNA-like precursor sequences, and many hairpin-shaped sequences which are not hairpin-shaped miRNA-like precursors.

[0224] In order to validate the performance of the HAIRPIN DETECTOR 114 (Fig. 9), preferably a variation of the k-fold cross validation method (Tom M. Mitchell, Machine Learning, McGraw Hill (1997)), is employed. This preferred validation method is well adapted to deal with the training set, which includes large numbers of similar or identical miRNAs. The training set is therefore preferably initially divided into groups of miRNAs such that any two miRNAs that belong to different groups have an Edit Distance score of at least D=3, i.e. they differ by at least 3 editing steps (Dan Gusfield, Algorithms on strings, trees, and sequences: computer science and computational biology, Cambridge University Press, 1997). Next, the groups are preferably classified into k sets of groups. Standard k-fold cross validation is preferably performed on these sets, preferably using k=5, such that the training set and the test set include at least one sequence from each of the groups. It is appreciated that without the prior grouping, standard cross validation methods incorrectly indicate much higher performance results for the predictors due to the redundancy of training examples within the genome of a species and across genomes of different species.

[0225] In a preferred embodiment of the present invention, using

the abovementioned validation methodology, the efficacy of the HAIRPIN DETECTOR 114 (Fig. 9) is confirmed. For example, when a similarity threshold is chosen such that 90% of the known hairpin-shaped miRNA precursors are successfully predicted, only 11% of the approximately 342,880 background set of hairpin-shaped sequences are predicted to be hairpin-shaped miRNA-like precursors.

[0226] Reference is now made to Fig. 12B which is a simplified flowchart illustrating preferred operation of the HAIRPIN DETECTOR 114 of Fig. 9. The HAIRPIN DETECTOR 114 preferably initially uses a secondary structure folding algorithm based on free-energy minimization, such as the MFOLD algorithm, described in Mathews et al. J. Mol. Biol. 288:911-940 (1999) and Zuker, M. Nucleic Acids Res. 31: 3406-3415. (2003), the disclosure of which is hereby incorporated by reference. This algorithm is operative to calculate probable secondary structure folding patterns of the NON-PROTEIN CODING GENOMIC SEQUENCES 136 (Fig. 11A) as well as the free-energy of each of these probable secondary folding patterns. The secondary structure folding algorithm, such as the MFOLD algorithm (Mathews, 1997; Zuker 2003), typically provides a listing of the base-pairing of the folded shape, i.e. a listing of

each pair of connected nucleotides in the sequence.

[0227] Next, the HAIRPIN DETECTOR 114 analyzes the results of the secondary structure folding patterns, in order to determine the presence and location of hairpin folding structures. The goal of this second step is to assess the base-pairing listing provided by the secondary structure folding algorithm, in order to determine whether the base-pairing listing describes one or more hairpin type bonding pattern. Preferably, sequence segment corresponding to a hairpin structure is then separately analyzed by the secondary structure folding algorithm in order to determine its exact folding pattern and free-energy.

[0228] The HAIRPIN DETECTOR 114 then assesses the hairpin structures found by the previous step, comparing them to hairpin structures of known miRNA precursors, using various characteristic hairpin structure features such as length of the hairpin structure, length of the loop of mismatched nucleotides at its center, its free-energy and its thermodynamic stability, the amount and type of mismatched nucleotides and the existence of sequence repeat-elements. Only hairpins that bear statistically significant resemblance to the training set of hairpin structures

of known miRNA precursors, according to the abovementioned parameters, are accepted.

[0229] In a preferred embodiment of the present invention, similarity to the training set of hairpin structures of known miRNA precursors is determined using a "similarity score" which is calculated using a weighted sum of terms, where each term is a function of one of the abovementioned hairpin structure features. The parameters of each function are learned from the set of hairpin structures of known miRNA precursors, as described hereinabove with reference to HAIRPIN DETECTOR TRAINING & VALIDATION FUNCTIONALITY 124 (Fig. 10). The weight of each term in the similarity score is optimized so as to achieve maximized separation between the distribution peaks of similarity scores validated miRNA-precursor hairpin structures, and the distribution of similarity scores of hairpin structures detected in the background set mentioned hereinabove with reference to Fig. 12B.

[0230] In an alternative preferred embodiment of the present invention, the step described in the preceding paragraph may be split into two stages. A first stage implements a simplified scoring method, typically based on thresholding a subset of the hairpin structure features described

hereinabove, and may employ a minimum threshold for hairpin structure length and a maximum threshold for free energy. A second stage is preferably more stringent, and preferably employs a full calculation of the weighted sum of terms described hereinabove. The second stage preferably is performed only on the subset of hairpin structures that survived the first stage.

[0231] The HAIRPIN DETECTOR 114 also attempts to select hairpin structures whose thermodynamic stability is similar to that of hairpin structures of known miRNA precursors. This may be achieved in various ways. A preferred embodiment of the present invention utilizes the following methodology, preferably comprising three logical steps:

[0232] First, the HAIRPIN DETECTOR 114 attempts to group hairpin structures into "families" of closely related hairpin structures. As is known in the art, a secondary structure folding algorithm typically provides multiple alternative folding patterns, for a given genomic sequence and indicates the free energy of each alternative folding pattern.. It is a particular feature of the present invention that the HAIRPIN DETECTOR 114 preferably assesses the various hairpin structures appearing in the various alternative folding patterns and groups hairpin structures which ap-

pear at identical or similar sequence locations in various alternative folding patterns into common sequence location based "families" of hairpins. For example, all hairpin structures whose center is within 7 nucleotides of each other may be grouped into a family". Hairpin structures may also be grouped into a family" if their nucleotide sequences are identical or overlap to a predetermined degree.

[0233] It is also a particular feature of the present invention that the hairpin structure "families" are assessed in order to select only those families which represent hairpin structures that are as thermodynamically stable as those of hairpin structures of known miRNA precursors. Preferably only families which are represented in at least a selected majority of the alternative secondary structure folding patterns, typically 65%, 80% or 100% are considered to be sufficiently stable.

[0234] It is an additional particular feature of the present invention that the most suitable hairpin structure is selected from each selected family. For example, a hairpin structure which has the greatest similarity to the hairpin structures appearing in alternative folding patterns of the family may be preferred. Alternatively or additionally, the

hairpin structures having relatively low free energy may be preferred.

[0235] Alternatively or additionally considerations of homology to hairpin structures of other organisms and the existence of clusters of thermodynamically stable hairpin structures located adjacent to each other along a sequence may be important in selection of hairpin structures. The tightness of the clusters in terms of their location and the occurrence of both homology and clusters may be of significance.

[0236] Reference is now made to Figs. 13A - 13C which together describe the structure and operation of the DICER-CUT LOCATION DETECTOR 116, described hereinabove with Fig. 9.

[0237] Fig. 13A is a simplified block diagram of a preferred implementation of the DICER-CUT LOCATION DETECTOR 116. The goal of the DICER-CUT LOCATION DETECTOR 116 is to detect the location in which the DICER COMPLEX, described hereinabove with reference to Fig. 8, dices GAM FOLDED PRECURSOR RNA, yielding GAM RNA.

[0238] The DICER-CUT LOCATION DETECTOR 116 therefore receives a plurality of HAIRPIN STRUCTURES ON GENOMIC SEQUENCES 138 (Fig. 12A), and, following operation of

DICER-CUT LOCATION DETECTOR TRAINING & VALIDATION FUNCTIONALITY 126 (Fig.10), is operative to detect a plurality of DICER-CUT SEQUENCES FROM HAIRPIN STRUCTURES 140.

[0239] Reference is now made to Fig. 13B which is a simplified flowchart illustrating a preferred implementation of DICER-CUT LOCATION DETECTOR TRAINING & VALIDATION FUNCTIONALITY 126.

[0240] A general goal of the DICER-CUT LOCATION DETECTOR TRAINING & VALIDATION FUNCTIONALITY 126 is to analyze the dicer-cut locations of known diced miRNA on respective hairpin shaped miRNA precursors in order to determine a common pattern in these locations, which can be used to predict dicer cut locations on GAM folded precursor RNAs.

[0241] The dicer-cut locations of known miRNA precursors are obtained and studied. Locations of the 5' and/or 3' ends of the known diced miRNAs are preferably represented by their respective distances from the 5' end of the corresponding hairpin shaped miRNA precursor. Additionally or alternatively, the 5' and/or 3' ends of the known diced miRNAs are preferably represented by the relationship between their locations and the locations of one or more nu-

cleotides along the hairpin shaped miRNA precursor. Additionally or alternatively, the 5' and/or 3' ends of the known diced miRNAs are preferably represented by the relationship between their locations and the locations of one or more bound nucleotide pairs along the hairpin shaped miRNA precursor. Additionally or alternatively, the 5' and/or 3' ends of the known diced miRNAs are preferably represented by the relationship between their locations and the locations of one or more mismatched nucleotide pairs along the hairpin shaped miRNA precursor. Additionally or alternatively, the 5' and/or 3' ends of the known diced miRNAs are preferably represented by the relationship between their locations and the locations of one or more unmatched nucleotides along the hairpin shaped miRNA precursor. Additionally or alternatively, locations of the 5' and/or 3' ends of the known diced miRNAs are preferably represented by their respective distances from the loop located at the center of the corresponding hairpin shaped miRNA precursor.

[0242] One or more of the foregoing location metrics may be employed in the training and validation functionality. Additionally, metrics related to the nucleotide content of the diced miRNA and/or of the hairpin shaped miRNA precur-

sor may be employed.

[0243] In a preferred embodiment of the present invention, DICER-CUT LOCATION DETECTOR TRAINING & VALIDATION FUNCTIONALITY 126 preferably employs standard machine learning techniques known in the art of machine learning for analysis of existing patterns in a given "training set" of examples. These techniques are capable, to a certain degree, of detecting similar patterns in other, previously unseen examples. Such machine learning techniques include, but are not limited to neural networks, Bayesian networks, Support Vector Machines (SVM), Genetic Algorithms, Markovian modeling, Maximum Likelihood modeling, Nearest Neighbor algorithms, Decision trees and other techniques, as is well known in the art.

[0244] In accordance with one embodiment of the present invention, machine learning predictors, such as a Support Vector Machine (SVM) predictor, are applied to the aforementioned training set and are operative, for example to test every possible nucleotide on a hairpin as a candidate for being the 5' end or the 3' end of a diced GAM RNA. More preferred machine learning predictors include predictors based on Nearest Neighbor, Bayesian modeling, and K-nearest-neighbor algorithms. A training set of the known

miRNA precursor sequences is preferably used for training multiple separate classifiers or predictors, each of which produces a model for the 5' and/or 3' end locations of a diced miRNA with respect to its hairpin precursor. The models take into account one or more of the various miRNA location metrics described above.

[0245] Performance of the resulting predictors, evaluated on the abovementioned validation set of 440 published miRNAs using k-fold cross validation (Mitchell, 1997) with k = 3, is found to be as follows: in 70% of known miRNAs 5'-end location is correctly determined by an SVM predictor within up to 2 nucleotides; a Nearest Neighbor (EDIT DISTANCE) predictor achieves 56% accuracy (247/440); a Two-Phased predictor that uses Bayesian modeling (TWO PHASED) achieves 80% accuracy (352/440), when only the first phase is used. When the second phase (strand choice) is implemented by a nave Bayesian model the accuracy is 55% (244/440), and when the K-nearest-neighbor modeling is used for the second phase, 374/440 decision are made and the accuracy is 65% (242/374). A K-nearest-neighbor predictor (FIRST-K) achieves 61% accuracy (268/440). The accuracies of all predictors are considerably higher on top scoring subsets of published miRNA.

[0246] Finally, in order to validate the efficacy and accuracy of the DICER-CUT LOCATION DETECTOR 116, a sample of novel oligonucleotides detected thereby is preferably selected, and validated by wet lab. Laboratory results validating the efficacy of the DICER-CUT LOCATION DETECTOR 116 are described hereinbelow with reference to Figs. 21 - 24D.

[0247] Reference is now made to Fig. 13C which is a simplified flowchart illustrating operation of DICER-CUT LOCATION DETECTOR 116 (Fig. 9), constructed and operative in accordance with a preferred embodiment of the present invention. The DICER CUT LOCATION DETECTOR 116 preferably comprises a machine learning computer program module, which is trained to recognize dicer-cut locations on known hairpin-shaped miRNA precursors, and based on this training, is operable to detect dicer-cut locations of novel GAM RNAs (Fig. 8) on GAM FOLDED PRECURSOR RNAs (Fig. 8). In a preferred embodiment of the present invention, the dicer-cut location module preferably utilizes machine learning algorithms, such as known Support Vector Machine (SVM) and more preferably: known Bayesian modeling, Nearest Neighbors, and K-nearest-neighbor algorithms.

[0248] When initially assessing a novel GAM FOLDED PRECURSOR RNA, all 19-24 nucleotide long segments thereof are initially considered as "potential GAM RNAs", since the dicer-cut location is initially unknown.

[0249] For each such potential GAM RNA, the location of its 5' end or the locations of its 5' and 3' ends are scored by at least one recognition classifier or predictor.

[0250] In a preferred embodiment of the present invention, the DICER-CUT LOCATION DETECTOR 116 (Fig. 9) may use a Support Vector Machine predictor trained on and operating on features such as the following:

[0251] Locations of the 5' and/or 3' ends of the known diced miRNAs, which are preferably represented by their respective distances from the 5' end of the corresponding hairpin shaped miRNA precursor. Additionally or alternatively, the 5' and/or 3' ends of the known diced miRNAs are preferably represented by the relationship between their locations and the locations of one or more nucleotides along the hairpin shaped miRNA precursor. Additionally or alternatively, the 5' and/or 3' ends of the known diced miRNAs are preferably represented by the relationship between their locations and the locations of one or more bound nucleotide pairs along the hairpin

shaped miRNA precursor. Additionally or alternatively, the 5' and/or 3' ends of the known diced miRNAs are preferably represented by the relationship between their locations and the locations of one or more mismatched nucleotide pairs along the hairpin shaped miRNA precursor. Additionally or alternatively, the 5' and/or 3' ends of the known diced miRNAs are preferably represented by the relationship between their locations and the locations of one or more unmatched nucleotides along the hairpin shaped miRNA precursor. Additionally or alternatively, locations of the 5' and/or 3' ends of the known diced miRNAs are preferably represented by their respective distances from the loop located at the center of the corresponding hairpin shaped miRNA precursor; and secondarily

[0252] Metrics related to the nucleotide content of the diced miRNA and/or of the hairpin shaped miRNA precursor.

[0253] In another preferred embodiment of the present invention, the DICER-CUT LOCATION DETECTOR 116 (Fig. 9) preferably employs an "EDIT DISTANCE" predictor, which seeks sequences that are similar to those of known miRNAs, utilizing a Nearest Neighbor algorithm, where a similarity metric between two sequences is a variant of the

Edit Distance algorithm (Gusfield, 1997). The EDIT DISTANCE predictor is based on an observation that miRNA oligonucleotides tend to form clusters , the members of which show marked sequence similarity.

[0254] In yet another preferred embodiment of the present invention, the DICER-CUT LOCATION DETECTOR 116 (Fig. 9) preferably uses a "TWO PHASE" predictor, which predicts the dicer-cut location in two distinct phases: (a) selecting a double-stranded segment of the GAM FOLDED PRECURSOR RNA (Fig. 8) comprising the GAM RNA by nave Bayesian modeling and (b) detecting which strand of the double-stranded segment contains GAM RNA (Fig. 8) by employing either nave or by K-nearest-neighbor modeling. K-nearest-neighbor modeling is a variant of the 'FIRST-K' predictor described hereinbelow, with parameters optimized for this specific task. The 'TWO PHASE' predictor may be operated in two modes: either utilizing only the first phase and thereby producing two alternative dicer-cut location predictions, or utilizing both phases and thereby producing only one final dicer-cut location.

[0255] In still another preferred embodiment of the present invention, the DICER-CUT LOCATION DETECTOR 116 preferably uses a "FIRST-K" predictor, which utilizes a K-

nearest-neighbor algorithm. The similarity metric between any two sequences is 1- E/L , where L is a parameter, preferably 8-10 and E is the edit distance between the two sequences, taking into account only the first L nucleotides of each sequence. If the K-nearest-neighbor scores of two or more locations on the GAM FOLDED PRECURSOR RNA (Fig. 8) are not significantly different, these locations are further ranked by a Bayesian model, similar to the one described hereinabove.

[0256] The TWO PHASE and FIRST-K predictors preferably are trained on and operate on features such as the following:

[0257] Locations of the 5' and/or 3' ends of the the known diced miRNAs, which are preferably represented by their respective distances from the 5' end of the corresponding hairpin shaped miRNA precursor. Additionally or alternatively, the 5' and/or 3' ends of the known diced miRNAs are preferably represented by the relationship between their locations and the locations of one or more nucleotides along the hairpin shaped miRNA precursor. Additionally or alternatively, the 5' and/or 3' ends of the known diced miRNAs are preferably represented by the relationship between their locations and the locations of one or more bound nucleotide pairs along the hairpin

shaped miRNA precursor. Additionally or alternatively, the 5' and/or 3' ends of the known diced miRNAs are preferably represented by the relationship between their locations and the locations of one or more mismatched nucleotide pairs along the hairpin shaped miRNA precursor. Additionally or alternatively, the 5' and/or 3' ends of the known diced miRNAs are preferably represented by the relationship between their locations and the locations of one or more unmatched nucleotides along the hairpin shaped miRNA precursor. Additionally or alternatively, locations of the 5' and/or 3' ends of the the known diced miRNAs are preferably represented by their respective distances from the loop located at the center of the corresponding hairpin shaped miRNA precursor; and secondarily

[0258] Metrics related to the nucleotide content of the diced miRNA and/or of the hairpin shaped miRNA precursor.

[0259] In accordance with an embodiment of the present invention scores of two or more of the abovementioned classifiers or predictors are integrated, yielding an integrated score for each "potential GAM RNA". As an example, Fig. 13C illustrates integration of scores from two classifiers, a 3' end recognition classifier and a 5' end recognition clas-

sifier, the scores of which are integrated to yield an integrated score. Most preferably, the INTEGRATED SCORE of Fig. 13C preferably implements a "best-of-breed" approach employing a pair of classifiers and accepting only "potential GAM RNAs" that score highly on one of the above mentioned "EDIT DISTANCE", or "TWO-PHASE" predictors. In this context, "high scores" means scores which have been demonstrated to have low false positive value when scoring known miRNA oligonucleotides. Alternatively, the INTEGRATED SCORE may be derived from operation of more or less than two classifiers.

[0260] The INTEGRATED SCORE is evaluated as follows: (a) the "potential GAM RNA" having the highest score is preferably taken to be the most probable GAM RNA, and (b) if the integrated score of this most probable GAM RNA is higher than a pre-defined threshold, then the most probable GAM RNA is accepted as a PREDICTED GAM RNA. Preferably, this evaluation technique is not limited to the highest scoring potential GAM RNA.

[0261] Reference is now made to Fig. 14A which is a simplified block diagram of a preferred implementation of the TARGET GENE BINDING-SITE DETECTOR 118 described hereinabove with reference to Fig. 9. The goal of the TARGET

GENE BINDING-SITE DETECTOR 118 is to detect one or more binding sites such as BINDING SITE I, BINDING SITE II and BINDING SITE III (Fig. 8) located in untranslated regions of the mRNA of a known gene, the nucleotide sequence of which binding sites is partially or fully complementary to a GAM RNA, thereby determining that the above mentioned known gene is a target gene thereof.

[0262] The TARGET GENE BINDING-SITE DETECTOR 118 (Fig. 9) receives a plurality of DICER-CUT SEQUENCES FROM HAIRPIN STRUCTURES 140 (Fig. 13A), and a plurality of POTENTIAL TARGET GENE SEQUENCES 142 which are derived from SEQUENCED DNA DATA 104 (Fig. 9).

[0263] TARGET GENE BINDING-SITE DETECTOR TRAINING & VALIDATION FUNCTIONALITY 128 (Fig.10) is operative to train the TARGET GENE BINDING-SITE DETECTOR on known miRNAs and their respective target genes. A sequence comparison of sequences of known miRNA oligonucleotides to sequences of known binding sites of known target thereof is performed by utilizing BLAST or other algorithms such as EDIT DISTANCE.

[0264] The results are preferably employed to define a threshold based on scoring distinctions between known miRNA binding sites and sequences which are known not to be

miRNA binding sites. This threshold is used during operation of TARGET GENE BINDING-SITE DETECTOR 118 to distinguish between miRNA-like binding sites of potential GAM RNA and other sequences.

[0265] Next, the binding sites are expanded, and determinations are made whether if nucleotide sequences immediately adjacent to the binding sites found by the sequence comparison algorithm (e.g. BLAST or EDIT DISTANCE), may improve the match. Free-energy and spatial structure are computed for the resulting binding sites. Binding sites which are clustered are strongly preferred and binding sites found in evolutionarily conserved sequences may also be preferred. Free energy, spatial structure and the above preferences are reflected in scoring.

[0266] The resulting scores, characteristic of known binding sites (e.g. binding sites of known miRNA oligonucleotides Lin-4 and Let-7 to target genes Lin-14, Lin-41, Lin 28 etc.), may be employed for detection of binding-sites of novel GAM RNAs.

[0267] Following operation of TARGET GENE BINDING-SITE DETECTOR TRAINING & VALIDATION FUNCTIONALITY 128 (Fig.10), the TARGET GENE BINDING-SITE DETECTOR 118 is operative to detect a plurality of POTENTIAL NOVEL

TARGET GENES HAVING BINDING-SITE/S 144 the nucleotide sequence of which is partially or fully complementary to that of each of the plurality of DICER-CUT SEQUENCES FROM HAIRPIN STRUCTURES 140. Preferred operation of the TARGET GENE BINDING-SITE DETECTOR 118 is further described hereinbelow with reference to Fig. 14B.

[0268] Reference is now made to Fig. 14B which is a simplified flowchart illustrating a preferred operation of the TARGET GENE BINDING-SITE DETECTOR 118 of Fig. 9. In a preferred embodiment of the present invention, the TARGET GENE BINDING-SITE DETECTOR 118 employs a sequence comparison algorithm such as BLAST in order to compare the nucleotide sequence of each of the plurality of DICER-CUT SEQUENCES FROM HAIRPIN STRUCTURES 140 (Fig. 13A), to the POTENTIAL TARGET GENE SEQUENCES 142 (Fig.14A), such as untranslated regions of known mRNAs, in order to find crude potential matches. Alternatively, the sequence comparison may be performed using a sequence match search tool that is essentially a variant of the EDIT DISTANCE algorithm described hereinabove with reference to Fig. 13C, and the Nearest Neighbor algorithm.

[0269] A sequence comparison of DICER-CUT SEQUENCES FROM HAIRPIN STRUCTURES 140 (Fig. 14A) are compared to POTENTIAL TARGET GENE SEQUENCES 142 (Fig. 14A) by utilizing BLAST or other algorithms such as EDIT DISTANCE.

[0270] The results are preferably filtered according to a threshold determined in accordance with the scoring resulting from the sequence comparison carried out by the TARGET GENE BINDING-SITE DETECTOR TRAINING & VALIDATION FUNCTIONALITY 128.

[0271] Next the binding sites are expanded, and determinations are made whether if nucleotide sequences immediately adjacent to the binding sites found by the sequence comparison algorithm (e.g. BLAST or EDIT DISTANCE), may improve the match.

[0272] Free-energy and spatial structure are computed for the resulting binding sites. Binding sites which are clustered are strongly preferred and binding sites found in evolutionarily conserved sequences may also be preferred. Free energy, spatial structure and the above preferences are reflected in scoring.

[0273] The resulting scores are compared with scores characteristic of known binding sites (e.g. binding sites of known miRNA oligonucleotides Lin-4 and Let-7 to target genes

Lin-14, Lin-41, Lin 28 etc.).

[0274] For each candidate binding site a score, here termed Binding Site Prediction Accuracy, is calculated which estimates its similarity to known binding sites. This score is based on GAM binding site characteristics including, but not limited to:

[0275] The free energy of binding of the GAM RNA - GAM RNA binding site complex;

[0276] Additionally or alternatively, the 5' and/or 3' ends of the GAM RNA, preferably represented by the relationship between their locations and the locations of one or more nucleotides along the GAM RNA; Additionally or alternatively, the 5' and/or 3' ends of the GAM RNA, preferably represented by the relationship between their locations and the locations of one or more bound nucleotide pairs along the GAM RNA binding site complex; Additionally or alternatively, the 5' and/or 3' ends of the GAM RNA, preferably represented by the relationship between their locations and the locations of one or more mismatched nucleotide pairs along the GAM RNA binding-site complex; Additionally or alternatively, the 5' and/or 3' ends of the GAM RNA, preferably represented by the relationship between their locations and the locations of one or more unmatched nu-

cleotides along the GAM RNA binding-site complex.

[0277] In accordance with another preferred embodiment of the present invention, binding sites are searched by a reversed process. Sequences of K (preferably 22) nucleotides of a untranslated regions of a target gene are assessed as potential binding sites. A sequence comparison algorithm, such as BLAST or EDIT DISTANCE, is then used to search elsewhere in the genome for partially or fully complementary sequences which are found in known miRNA oligonucleotides or computationally predicted GAM oligonucleotides. Only complementary sequences, which meet predetermined spatial structure and free energy criteria as described hereinabove are accepted. Clustered binding sites are strongly preferred and potential binding sites and potential GAM oligonucleotides which occur in evolutionarily conserved genomic sequences are also preferred. Scoring of candidate binding sites takes into account free energy and spatial structure of the binding site complexes, as well as the aforesaid preferences.

[0278] Target binding sites identified by the TARGET GENE BINDING-SITE DETECTOR 118 (Fig. 9), are preferably divided into four groups:

[0279] a) binding sites which are exactly complementary to the

predicted GAM RNA. (1nt. mismatch is allowed)

[0280] b) binding sites which are not exactly complementary to the predicted GAM RNA and having 0.8=< Binding Site Prediction Accuracy < 1;

[0281] c) binding sites which are not exactly complementary to the predicted GAM RNA and having 0.7= < Binding Site Prediction Accuracy < 0.8; and

[0282] d) binding sites which are not exactly complementary to the predicted GAM RNA and having 0.6= < Binding Site Prediction Accuracy < 0.7.

[0283] The average number of mismatched nucleotides in the alignment of predicted GAM RNA and a corresponding target gene binding-site is smallest in category a and largest in category d.

[0284] In accordance with a preferred embodiment of the present invention there is provided a binding site specific ranking, indicative of the degree of similarity of characteristics of the binding of a GAM to a target gene binding site, to binding characteristic of known miRNAs. This ranking preferably utilizes the evaluation criteria described here-inabove.

[0285] In accordance with another preferred embodiment of the present invention, there is provided a UTR specific ranking

of GAM to target gene binding. , indicative of the degree of similarity of characteristics of the binding of a GAM to a cluster of target gene binding sites on a UTR, to binding characteristics of known miRNAs to UTRs of corresponding miRNA target genes. This ranking preferably is a weighted sum of the binding site specific rankings of various clustered binding sites.

[0286] Reference is now made to Fig. 15 which is a simplified flowchart illustrating a preferred operation of the FUNCTION & UTILITY ANALYZER 120 described hereinabove with reference to Fig. 9. The goal of the FUNCTION & UTILITY ANALYZER 120 is to determine if a potential target gene is in fact a valid clinically useful target gene. Since a potential novel GAM oligonucleotide binding a binding site in the UTR of a target gene is understood to inhibit expression of that target gene, and if that target gene is shown to have a valid clinical utility, then in such a case it follows that the potential novel oligonucleotide itself also has a valid useful function which is the opposite of that of the target gene.

[0287] The FUNCTION & UTILITY ANALYZER 120 preferably receives as input a plurality of POTENTIAL NOVEL TARGET GENES HAVING BINDING-SITE/S 144 (Fig. 14A), generated

by the TARGET GENE BINDING-SITE DETECTOR 118 (Fig. 9). Each potential oligonucleotide is evaluated as follows: First, the system checks to see if the function of the potential target gene is scientifically well established. Preferably, this can be achieved bioinformatically by searching various published data sources presenting information on known function of proteins. Many such data sources exist and are published as is well known in the art. Next, for those target genes the function of which is scientifically known and is well documented, the system then checks if scientific research data exists which links them to known diseases. For example, a preferred embodiment of the present invention utilizes the OMIM(TM) (Hamosh et al, 2002) database published by NCBI, which summarizes research publications relating to genes which have been shown to be associated with diseases. Finally, the specific possible utility of the target gene is evaluated. While this process too may be facilitated by bioinformatic means, it might require manual evaluation of published scientific research regarding the target gene, in order to determine the utility of the target gene to the diagnosis and or treatment of specific disease. Only potential novel oligonucleotides, the target genes of which have passed all three

examinations, are accepted as novel oligonucleotide.

[0288] Reference is now made to Fig. 16, which is a simplified diagram describing each of a plurality of novel bioinformatically detected regulatory polynucleotide, referred to here as Genomic Record (GR) polynucleotide which encodes an `operon-like` cluster of novel micro RNA-like oligonucleotides each of which in turn modulates expression of at least one target gene, the function and utility of which at least one target gene is known in the art. GR GPRECURSOR DNA is a novel bioinformatically detected regulatory, non protein coding, polynucleotide. The method by which GR polynucleotide as detected is described hereinabove with additional reference to Figs. 9-18. GR GPRECURSOR DNA encodes GR PRECURSOR RNA, an RNA molecule, typically several hundreds to several thousands nucleotides long. GR PRECURSOR RNA folds spatially, forming GR FOLDED PRECURSOR RNA. It is appreciated that GR FOLDED PRECURSOR RNA comprises a plurality of what is known in the art as `hairpin` structures. These `hairpin` structures are due to the fact that the nucleotide sequence of GR PRECURSOR RNA comprises a plurality of segments, the first half of each such segment having a nucleotide sequence which is at least a partial or accurate complementary se-

quence of the second half thereof, as is well known in the art. GR FOLDED PRECURSOR RNA is naturally processed by cellular enzymatic activity into separate GAM precursor RNAs, herein schematically represented by GAM1 FOLDED PRECURSOR RNA through GAM3 FOLDED PRECURSOR RNA, each of which GAM precursor RNAs being a hairpin shaped RNA segment, corresponding to GAM FOLDED PRECURSOR RNA of Fig.8. The above mentioned GAM precursor RNAs are diced by DICER COMPLEX of Fig.8, yielding respective short RNA segments of about 22 nucleotides in length, schematically represented by GAM1 RNA through GAM3 RNA, each of which GAM RNAs corresponding to GAM RNA of Fig. 8. GAM1 RNA, GAM2 RNA and GAM3 RNA, each bind complementarily to binding sites located in untranslated regions of respective target genes, designated GAM1-TARGET RNA, GAM2-TARGET RNA and GAM3-TARGET RNA, respectively, which target binding site corresponds to a target binding site such as BINDING SITE I, BINDING SITE II or BINDING SITE III of Fig.8. This binding inhibits translation of the respective target proteins designated GAM1-TARGET PROTEIN, GAM2-TARGET PROTEIN and GAM3-TARGET PROTEIN respectively. It is appreciated that specific functions, and

accordingly utilities, of each GR polynucleotides of the present invention, correlates with, and may be deduced from, the identity of the target genes, which are inhibited by GAM RNAs comprised in the `operon-like` cluster of said GR polynucleotide schematically represented by GAM1 TARGET PROTEIN through GAM3 TARGET PROTEIN.

[0289] A listing of GAM oligonucleotide comprised in each of a plurality of GR polynucleotide of Fig. 16 is provided in Table 10, hereby incorporated herein. Nucleotide sequences of each said GAM oligonucleotide and their respective genomic source and chromosomal location are further described hereinbelow with reference to Table 3 hereby incorporated herein. GAM TARGET GENEs of each of said GAM oligonucleotides are elaborated hereinbelow with reference to Table 7, hereby incorporated herein. The functions of each of said GAM TARGET GENEs and their association with various diseases, and accordingly the utilities of said each of GAM oligonucleotides and hence the functions and utilities of each of said GR polynucleotides are elaborated hereinbelow with reference to Table 8, hereby incorporated herein. Studies establishing known functions of each of said GAM TARGET GENEs, and correlation of each of said GAM TARGET GENEs to known

diseases are listed in Table 9, and are hereby incorporated herein.

[0290] The present invention discloses 313 novel genes of the GR group of polynucleotides, which have been detected bioinformatically. Laboratory confirmation of 2 polynucleotides of the GR group of polynucleotides is described hereinbelow with reference to Figs. 23A - 24D.

[0291] In summary, the current invention discloses a very large number of novel GR polynucleotides each of which encodes a plurality of GAM oligonucleotides, which in turn may modulate expression of a plurality of target proteins. It is appreciated therefore that the function of GR polynucleotides is in fact similar to that of the Genomic Records concept of the present invention addressing the differentiation enigma, described hereinabove with reference to Fig. 7.

[0292] Reference is now made to Fig. 17 which is a simplified diagram illustrating a mode by which oligonucleotides of a novel group of operon-like polynucleotide described hereinabove with reference to Fig. 16 of the present invention, modulate expression of other such polynucleotide, in a cascading manner.GR1 PRECURSOR DNA and GR2 PRECURSOR DNA are two polynucleotides of the novel

group of operon-like polynucleotides designated GR PRECURSOR DNA(Fig. 16). As is typical of polynucleotides of the GR group of polynucleotides GR1 PRECURSOR DNA and GR2 PRECURSOR DNA, each encode a long RNA precursor, which in turn folds into a folded RNA precursor comprising multiple hairpin shapes, and is cut into respective separate hairpin shaped RNA segments, each of which RNA segments being diced to yield a n oligonucleotide of a group of oligonucleotide designated GAM RNA. In this manner GR1 yields GAM1 RNA, GAM2 RNA and GAM3 RNA, and GR2 yields GAM4 RNA, GAM5 RNA and GAM6 RNA. As Fig. 17 shows, GAM3 RNA , which derives from GR1, binds a binding site located adjacent to GR2 GPRECURSOR DNA thus modulating expression of GR2, thereby invoking expression of GAM4 RNA , GAM5 RNA and GAM6 RNA which derive from GR2. It is appreciated that the mode of modulation of expression presented by Fig. 17 enables an unlimited `cascading effect` in which a GR polynucleotide comprises multiple GAM oligonucleotides each of which may modulate expression of other GR polynucleotides each such GR polynucleotides comprising additional GAM oligonucleotide etc., whereby eventually certain GAM oligonucleotides modulate expres-

sion of target proteins. This mechanism is in accord with the conceptual model of the present invention addressing the differentiation enigma, described hereinabove with specific reference to Figs. 6 - 7.

[0293] Reference is now made to Fig. 18 which is a block diagram illustrating an overview of a methodology for finding novel oligonucleotides and operon-like polynucleotides of the present invention, and their respective functions. According to a preferred embodiment of the present invention, the methodology to finding novel oligonucleotides of the present invention and their function comprises of the following major steps: First, FIND GAM OLIGONUCLEOTIDES 146 is used to detect , oligonucleotide of the novel group of oligonucleotide of the present invention, referred to here as GAM oligonucleotide. GAM oligonucleotides are located and their function elicited by detecting target proteins they bind and the function of those target proteins, as described hereinabove with reference to Figs. 9 - 15. Next, FIND GR POLYNUCLEOTIDES 147 is used to detect polynucleotide of a novel group of operon-like polynucleotide of the present invention, referred to here as GR polynucleotide. GR polynucleotides are located, by locating clusters of proximally located GAM oligonucleotide,

based on the previous step. Consequently, FIND HIERARCHY OF GR POLYNUCLEOTIDES 148 elicits the hierarchy of GR and GAM: binding sites for non-protein-binding GAM oligonucleotide comprised in each GR polynucleotide found are sought adjacent to other GR polynucleotides. When found, such a binding site indicates that the connection between the GAM and the GR the expression of which it modulates, and thus the hierarchy of the GR polynucleotides and the GAM oligonucleotides they comprise. Lastly, DEDUCE FUNCTION OF HIGH GR POLYNUCLEOTIDES AND GAM OLIGONUCLEOTIDES 149 is used to deduce the function of GR polynucleotides and GAM oligonucleotides which are `high` in the hierarchy, i.e. GAM oligonucleotides which modulate expression of other GR polynucleotides rather than directly modulating expression of target proteins. A preferred approach is as follows: The function of protein-modulating GAM oligonucleotides is deducible from the proteins which they modulate, provided that the function of these target proteins is known. The function of `higher` GAM oligonucleotides may be deduced by comparing the function of protein-modulating GAM oligonucleotides with the hierarchical relationships by which the `higher` GAM oligonu-

cleotides are connected to the protein-modulating GAM oligonucleotides. For example, given a group of several protein-modulating GAM oligonucleotides which collectively cause a protein expression pattern typical of a certain cell-type, then a `higher` GAM oligonucleotide is sought which modulates expression of GR polynucleotides which perhaps modulate expression of other GR polynucleotides which eventually modulate expression of the given group of protein- modulating GAM oligonucleotide. The `higher` GAM oligonucleotide found in this manner is taken to be responsible for differentiation of that cell-type, as per the conceptual model of the invention described hereinabove with reference to Fig. 6.

[0294] Reference is now made to Fig. 19 which is a block diagram illustrating different utilities of oligonucleotide of the novel group of oligonucleotides of the present invention referred to here as GAM oligonucleotides and GR polynucleotides. The present invention discloses a first plurality of novel oligonucleotides referred to here as GAM oligonucleotides and a second plurality of operon-like polynucleotides referred to here as GR polynucleotides each of the GR polynucleotide encoding a plurality of GAM oligonucleotides The present invention further discloses a

very large number of known target genes, which are bound by, and the expression of which is modulated by each of the novel oligonucleotides of the present invention. Published scientific data referenced by the present invention provides specific, substantial, and credible evidence that the above mentioned target genes modulated by novel oligonucleotides of the present invention, are associated with various diseases. Specific novel oligonucleotides of the present invention, target genes thereof and diseases associated therewith, are described hereinbelow with reference to Tables 1 through 11. It is therefore appreciated that a function of GAM oligonucleotides and GR polynucleotides of the present invention is modulation of expression of target genes related to known diseases, and that therefore utilities of novel oligonucleotides of the present invention include diagnosis and treatment of the above mentioned diseases. Fig. 19 describes various types of diagnostic and therapeutic utilities of novel oligonucleotides of the present invention. A utility of novel oligonucleotides of the present invention is detection of GAM oligonucleotides and of GR polynucleotides. It is appreciated that since GAM oligonucleotides and polynucleotides modulate expression of dis-

ease related target genes, that detection of expression of GAM oligonucleotides in clinical scenarios associated with said diseases is a specific, substantial and credible utility. Diagnosis of novel oligonucleotides of the present invention may preferably be implemented by RNA expression detection techniques, including but not limited to biochips, as is well known in the art. Diagnosis of expression of oligonucleotides of the present invention may be useful for research purposes, in order to further understand the connection between the novel oligonucleotides of the present invention and the above mentioned related diseases, for disease diagnosis and prevention purposes, and for monitoring disease progress. Another utility of novel oligonucleotides of the present invention is anti-GAM therapy, a mode of therapy which allows up regulation of a disease related target gene of a novel GAM oligonucleotide of the present invention, by lowering levels of the novel GAM oligonucleotide which naturally inhibits expression of that target gene. This mode of therapy is particularly useful with respect to target genes which have been shown to be under-expressed in association with a specific disease. Anti-GAM therapy is further discussed hereinbelow with reference to Figs. 20A and

20B. A further utility of novel oligonucleotides of the present invention is GAM replacement therapy, a mode of therapy which achieves down regulation of a disease related target gene of a novel GAM oligonucleotide of the present invention, by raising levels of the GAM which naturally inhibits expression of that target gene. This mode of therapy is particularly useful with respect to target genes which have been shown to be over-expressed in association with a specific disease. GAM replacement therapy involves introduction of supplementary GAM products into a cell, or stimulation of a cell to produce excess GAM products. GAM replacement therapy may preferably be achieved by transfecting cells with an artificial DNA molecule encoding a GAM which causes the cells to produce the GAM product, as is well known in the art. Yet a further utility of novel oligonucleotides of the present invention is modified GAM therapy. Disease conditions are likely to exist, in which a mutation in a binding site of a GAM RNA prevents natural GAM RNA to effectively bind inhibit a disease related target gene, causing up regulation of that target gene, and thereby contributing to the disease pathology. In such conditions, a modified GAM oligonucleotides is designed which effectively

binds the mutated GAM binding site, i.e. is an effective anti-sense of the mutated GAM binding site, and is introduced in disease effected cells. Modified GAM therapy is preferably achieved by transfecting cells with an artificial DNA molecule encoding the modified GAM which causes the cells to produce the modified GAM product, as is well known in the art. An additional utility of novel GAM of the present invention is induced cellular differentiation therapy. As aspect of the present invention is finding oligonucleotides which determine cellular differentiation, as described hereinabove with reference to Fig. 18. Induced cellular differentiation therapy comprises transfection of cell with such GAM oligonucleotides thereby determining their differentiation as desired. It is appreciated that this approach may be widely applicable, inter alia as a means for auto transplantation harvesting cells of one cell-type from a patient, modifying their differentiation as desired, and then transplanting them back into the patient. It is further appreciated that this approach may also be utilized to modify cell differentiation in vivo, by transfecting cells in a genetically diseased tissue with a cell-differentiation determining GAM thus stimulating these cells to differentiate appropriately.

[0295] Reference is now made to Figs. 20A and 20B, simplified diagrams which when taken together illustrate anti-GAM therapy mentioned hereinabove with reference to Fig. 19. A utility of novel GAMs of the present invention is anti-GAM therapy, a mode of therapy which allows up regulation of a disease related target gene of a novel GAM of the present invention, by lowering levels of the novel GAM which naturally inhibits expression of that target gene. Fig. 20A shows a normal GAM inhibiting translation of a target gene of GAM RNA by binding to a BINDING SITE found in an untranslated region of GAM TARGET RNA, as described hereinabove with reference to Fig. 8.

[0296] Fig. 20B shows an example of anti-GAM therapy. ANTI-GAM RNA is short artificial RNA molecule the sequence of which is an anti-sense of GAM RNA. Anti-GAM treatment comprises transfecting diseased cells with ANTI-GAM RNA, or with a DNA encoding thereof. The ANTI-GAM RNA binds the natural GAM RNA, thereby preventing binding of natural GAM RNA to its BINDING SITE. This prevents natural translation inhibition of GAM TARGET RNA by GAM RNA, thereby up regulating expression of GAM TARGET PROTEIN.

[0297] It is appreciated that anti-GAM therapy is particularly use-

ful with respect to target genes which have been shown to be under-expressed in association with a specific disease. Furthermore, anti-GAM therapy is particularly useful, since it may be used in situations in which technologies known in the art as RNAi and siRNA can not be utilized. As in known in the art, RNAi and siRNA are technologies which offer means for artificially inhibiting expression of a target protein, by artificially designed short RNA segments which bind complementarily to mRNA of said target protein. However, RNAi and siRNA can not be used to directly up regulate translation of target proteins.

[0298] Reference is now made to Fig. 21A, which is a histogram representing the distribution of known miRNA oligonucleotide s and that of hairpin structures extracted from expressed genome sequences with respect to their HAIRPIN DETECTOR score. The known miRNA oligonucleotide s set is taken from RFAM database, Release2.1 and include 440 miRNA oligonucleotide s from H.sapienas, M.musculus, C.elegans ,C.brigassae and D.melanogaster. Folding of expressed genome sequences taken from public databases of ESTs (Unigene-NCBI and TIGR) identified 342,882 hairpin structures.~154,000 out of the 342,882 hairpin structures did not pass the filter of being identi-

fied as hairpins in several secondary structure folding versions of the given genomic sequence , as described hereinabove with reference to Fig.12B, and hence did not receive a Hairpin detector score. Furthermore, ~133,000 hairpin structures did not pass the filter of minimum score of the DICER-CUT LOCATION DETECTOR 116 (Fig. 9) (those ~287,000 hairpin structures are not represented in the histogram). Hairpin structures are considered as miRNA -like precursor oligonucleotides here referred to as GAM oligonucleotide, if their Hairpin detector score is above 0.3. Thus, the GAM oligonucleotides set is comprised of ~40,000 hairpin structures, of those ~5100 received a high Hairpin detector score (>=0.7). These are much higher numbers than those of the known miRNA oligonucleotide s and of the upper bound of ~255 human miRNA oligonucleotide s, estimated by Bartel et al (Science,299,1540, March 2003). Of the reference set that pass the above filter (408/440), 284 (69%) received a high Hairpin detector score (>=0.7).

[0299] Reference is now made to Fig. 21B, which is a table summarizing laboratory validation results that validate efficacy of the BIOINFORMATIC OLIGONUCLEOTIDE DETECTION ENGINE 100 (Fig. 9). In order to assess efficacy of the

BIOINFORMATIC OLIGONUCLEOTIDE DETECTION ENGINE 100, novel oligonucleotides predicted thereby are preferably divided into 4 detection accuracy groups (first column), designated A through D, ranking GAMS from the most probable GAMs to the least probable GAMs, using the scores of HAIRPIN DETECTOR 114 (Fig. 9) and DICER-CUT LOCATION DETECTOR 116 (Fig. 9) as follows:

[0300] Group A: The score of the HAIRPIN-DETECTOR is above 0.7, the overall score of the two-phased predictor is above 0.55, and the score of the second phase of the two-phased predictor is above 0.75, or the score of the EDIT-DISTANCE predictor is equal or above 17. In this group, one Dicer cut location is predicted for each hairpin. Group B: The score of the HAIRPIN-DETECTOR is above 0.5, the overall score of the two-phased predictor is above 0.55, and the hairpin is not in group A. Group C: The score of the HAIRPIN-DETECTOR is between 0.4 and 0.5, and the overall score of the two-phased predictor is above 0.55. Group D: The score of the HAIRPIN-DETECTOR is between 0.3 and 0.4, and the overall score of the two-phased predictor is above 0.55. In groups B, C and D, if the score of the second phase of the two-phased predictor is above 0.75, one Dicer cut location is predicted

for each hairpin, otherwise both sides of the double stranded window are given as output, and are examined in the lab or used for binding site search. The groups are mutually exclusive, i.e. in groups A, C and D all hairpins score less than 17 in the EDIT-DISTANCE predictor.

[0301] It is appreciated that the division into groups is not exhaustive: 410 of the 440 published hairpins (second column), and 1891 of the 2147 novel GAMs, belong to one of the groups. An indication of the real performance of the two-phased predictor in the presence of background hairpins is given by the column 'precision on hairpin mixture' (third column). The precision on hairpin mixture is computed by mixing the published miRNA hairpins with background hairpins in a ratio of 1:4 and taking as a working assumption that they are hairpins not carrying a 'diced' miRNA.-like oligonucleotide This is a strict assumption, since some of these background hairpins may indeed contain 'diced' miRNAs-like oligonucleotide, while in this column they are all counted as failures

[0302] Sample novel bioinformatically predicted human GAMs of each of these groups are sent to the laboratory for validation (fourth column), and the number (fifth column) and percent (sixth column) of successful validation of pre-

dicted human GAM is noted for each of the groups, as well as overall (bottom line). The number of novel VAM genes explicitly specified by present invention belonging to each of the four groups is noted (seventh column).

[0303] It is appreciated that the present invention comprises 1891 novel GAM oligonucleotides, which fall into one of these four detection accuracy groups, and that the BIOINFORMATIC OLIGONUCLEOTIDE DETECTION ENGINE 100 (Fig. 9) is substantiated by a group of 52 novel human GAM oligonucleotides validated by laboratory means, out of 168 human GAM oligonucleotides which were tested in the lab, resulting in validation of an overall 31% accuracy. The top group demonstrated 37% accuracy. Pictures of test-results of specific human GAM oligonucleotides in the abovementioned four groups, as well as the methodology used for validating the expression of predicted oligonucleotides are elaborated hereinbelow with reference to Fig. 22.

[0304] It is further appreciated that failure to detect a predicted GAM oligonucleotide in the lab does not necessarily indicate a mistaken bioinformatic prediction. Rather, it may be due to technical sensitivity limitation of the lab test, or because the GAM oligonucleotide s not expressed in the

tissue examined, or at the development phase tested.

[0305] It is still further appreciated that in general these findings are in agreement with the expected bioinformatic accuracy, as describe hereinabove with reference to Fig. 13B: assuming 80% accuracy of the HAIRPIN DETECTOR 114 and 80% accuracy of the DICER-CUT LOCATION DETECTOR 116 and 80% accuracy of the lab validation, this would result in 50% overall accuracy of the GAM oligonucleotide validated in the lab.

[0306] Reference is now made to Fig. 22A which is a picture of laboratory results validating the expression of 43 novel genes detected by the BIOINFORMATIC OLIGONUCLEOTIDE DETECTION ENGINE 100 (Fig. 9).

[0307] Reference is now made to Fig. 22A and Fig.22B which are pictures and a summary table of laboratory results validating the expression of 43 novel human GAM oligonucleotides detected by the BIOINFORMATIC OLIGONUCLEOTIDE DETECTION ENGINE100. In each row in Fig.22A, pictures of several oligonucleotides validated by hybridization of Polymerase Chain Reaction (PCR)-product southern-blots, are provided, each corresponding to a specific GAM oligonucleotides, as elaborated hereinbelow. To test our validation method, we used a reference set of

8 known human miRNA oligonucleotide s , as blind test to our laboratory. These PCR-product hybridization pictures are designated 1 through 8 for the reference set known miRNA oligonucleotide s; and 9 through 51 for predicted GAM oligonucleotides.

[0308] In each PCR hybridization picture, 2 lanes are seen: the test lane, designated "+" and the control lane, designated "-". For convenience of viewing the results, all PCR-product hybridization pictures of Fig. 22A have been shrunk x4 vertically. It is appreciated that for each of the tested GAM oligonucleotides a clear hybridization band appears in the test ("+") lane, but not in the control ("-") lane.

[0309] Specifically, Fig. 22A shows pictures of PCR-product hybridization validation by southern-blot, the methodology of which is described hereinbelow, to the following novel human GAM oligonucleotides (RosettaGenomics Ltd. Nomenclature, 'A' and 'B' referred to the Dicer Cut Location as described hereinabelow with reference to the description of large tables:

[0310] (1)hsa-MIR-21;(2)hsa-MIR-27b; (3)hsa-MIR-186; (4)hsa-MIR-93; (5)hsa-MIR-26a ; (6)hsa-MIR-191; (7)hsa-MIR-31; (8)hsa-MIR-92; (9) GAM3418-A (later

published by other researchers as hsa-MIR23); (10) GAM4426-A; (11) GAM281-A; (12) GAM7553-A; (13) GAM5385-A; (14) GAM2608-A; (15) GAM1032-A; (16) GAM3431-A; (17) GAM7933-A; (18) GAM3298-A.; (19) GAM7080-A; (20)GAM895-A.; (21) GAM3770.1; (22) GAM337162-A; (23) GAM 8678-A; (24) GAM2033-A; (25)GAM7776-A; (26) GAM8145-A; (27) GAM25-A; (28) GAM7352.1; (29) GAM337624-A; (30) GAM1479-A; (31) GAM2270-A; (32) GAM7591-A; (33) GAM8285-A; (34) GAM6773-A; (35) GAM336818-A; (36) GAM336487-A; (37) GAM337620-A; (38) GAM336809-A; (39) GAM5346-A; (40) GAM8554-A; (41)GAM2071-A; (42) GAM7957-A; (43) GAM391-A; (44) GAM6633-A; (45) GAM19; (46) GAM8358-A; (47) GAM3229-A; an)GAM 7052-A; (49) GAM3027-A; (50) GAM21and (51) GAM oligonucleotide similar to mmu-MIR-30e.

[0311] The next validated GAM oligonucleotides are highly similar or highly identical to known mouse-miRNA oligonucleotides: GAM3027-A, similar to mmu-MIR-29c; GAM21, similar to mmu-MIR-130b; and GAM oligonucleotide which is highly similar to mmu-MIR-30e (picture number 51). In addition to the PCR- product hybridization detection, the following GAMs were cloned and sequenced :

GAM3418-A , GAM5385-A, GAM1032-A, GAM3298-A, GAM7080-A, GAM1338-A, GAM7776-A, GAM25-A, GAM337624-A, GAM1479-A, GAM6773-A, GAM336818-A, GAM336487-A, GAM337620-A, GAM336809-A, GAM3027-A, GAM21, and GAM oligonucleotide similar to mmu-MIR-30e (picture number 51). Furthermore, the following GAM oligonucleotides were sequenced directly from the ligation reaction by the method described hereinbelow under LIGATION-PCR DIAGNOSTIC METHOD: GAM4426-A, GAM7553-A, GAM2270-A, and GAM7591-A.

[0312] In order to validate the expression of predicted novel GAM and assuming that these novel GAM oligonucleotides are probably expressed at low concentrations, a PCR product cloning approach was set up through the following strategy: two types of cDNA libraries designated "One tailed" and "Ligation" were prepared from frozen HeLa S100 extract (4c Biotech, Belgium) size fractionated RNA. Essentially, Total S100 RNA was prepared through an SDSProteinase K incubation followed by an acid Phenol-Chloroform purification and Isopropanol precipitation. Alternatively, total HeLa RNA was also used as starting material for these libraries.

[0313] Fractionation was done by loading up to 500g per YM100 Amicon Microcon column (Millipore) followed by a 500g centrifugation for 40 minutes at 4C. Flow through "YM100"RNA consisting of about of the total RNA was used for library preparation or fractionated further by loading onto a YM30 Amicon Microcon column (Millipore) followed by a 13,500g centrifugation for 25 minutes at 4C. Flowthrough "YM30" was used for library preparation as is and consists of less than 0.5% of total RNA. For the both the "ligation" and the "One-tailed" libraries, RNA was dephosphorilated and ligated to an RNA (lowercase)-DNA (UPPERCASE) hybrid 5"-phosphorilated, 3" idT blocked 3"-adapter (5"-P-uuuAACCGCATCCTTCTC-idT-3" Dharmacon # P-002045-01-05) (as elaborated in Elbashir et al., Genes Dev. 15:188-200 (2001)) resulting in ligation only of RNase III type cleavage products. 3"-Ligated RNA was excised and purified from a half 6%, half 13% polyacrylamide gel to remove excess adapter with a Nanosep 0.2M centrifugal device (Pall) according to instructions, and precipitated with glycogen and 3 volumes of Ethanol. Pellet was resuspended in a minimal volume of water.

[0314] For the "ligation" library a DNA (UPPERCASE)-RNA (lowercase) hybrid 5"-adapter

(5"-TACTAATACGACTCACTaaa-3" Dharmacon # P-002046-01-05) was ligated to the 3"-adapted RNA, reverse transcribed with "EcoRI-RT": (5"-GACTAGCTGGAATTCAAGGATGCGGTTAAA-3"), PCR amplified with two external primers essentially as in Elbashir et al 2001 except that primers were "EcoRI-RT"and "PstI Fwd"(5"-CAGCCAACGCTGCAGATACGACTCACTAAA-3"). This PCR product was used as a template for a second round of PCR with one hemispecific and one external primer or with two hemispecific primers.

[0315] For the "One tailed" library the 3"-Adapted RNA was annealed to 20pmol primer "EcoRI RT" by heating to 70C and cooling 0.1C/sec to 30C and then reverse transcribed with Superscript II RT (According to instructions, Invitrogen) in a 20l volume for 10 alternating 5 minute cycles of 37C and 45C. Subsequently, RNA was digested with 1l 2M NaOH, 2mM EDTA at 65C for 10 minutes. cDNA was loaded on a polyacrylamide gel, excised and gel-purified from excess primer as above (invisible, judged by primer run alongside) and resuspended in 13l of water. Purified cDNA was then oligo-dC tailed with 400U of recombinant terminal transferase (Roche molecular biochemicals), 1l

100M dCTP, 1l 15mM CoCl2, and 4l reaction buffer, to a final volume of 20l for 15 minutes at 37C. Reaction was stopped with 2l 0.2M EDTA and 15l 3M NaOAc pH 5.2. Volume was adjusted to 150l with water, Phenol : Bromochloropropane 10:1 extracted and subsequently precipitated with glycogen and 3 volumes of Ethanol. C-tailed cDNA was used as a template for PCR with the external primers "T3-PstBsg(G/I)18"(5"-AATTAACCCTCACTAAAGGCTGCAG GTGCAGGIGGGIIGGGIIGGGIIGN-3" where I stands for Inosine and N for any of the 4 possible deoxynucleotides), and with "EcoRI Nested"(5"-GGAATTCAAGGATGCGGTTA-3"). This PCR product was used as a template for a second round of PCR with one hemispecific and one external primer or with two hemispecific primers.

[0316] Hemispecific primers were constructed for each predicted GAM RNA oligonucleotide by an in-house program designed to choose about half of the 5" or 3" sequence of the GAM RNA corresponding to a TM of about 30-34C constrained by an optimized 3" clamp, appended to the cloning adapter sequence (for "One-tailed" libraries 5"-GGNNGGGNNG on the 5" end of the GAM RNA , or TT-

TAACCGCATC-3" on the 3" end of the GAM RNA. For "Ligation" libraries the same 3" adapter and 5"-CGACTCACTAAA on the 5" end). Consequently, a fully complementary primer of a TM higher than 60C was created covering only one half of the GAM RNA sequence permitting the unbiased elucidation by sequencing of the other half.

[0317] CONFIRMATION OF GAM OLIGONUCLEOTIDE SEQUENCE AUTHENTICITY OF PCR PRODUCTS:

[0318] SOUTHERN BLOT:PCR-product sequences were confirmed by southern blot (Southern E.M., Biotechnology, 1992,24:122-39 (1975)) and hybridization with DNA oligonucleotide probes synthesized against predicted GAM RNAs oligonucleotides. Gels were transferred onto a Biodyne PLUS 0.45m, (Pall) positively charged nylon membrane and UV cross-linked. Hybridization was performed overnight with DIG-labeled probes at 420C in DIG Easy-Hyb buffer (Roche). Membranes were washed twice with 2xSSC and 0.1% SDS for 10 min. at 420C and then washed twice with 0.5xSSC and 0.1% SDS for 5 min at 420C. The membrane was then developed by using a DIG luminescent detection kit (Roche) using anti-DIG and CSPD reaction, according to the manufacturer's protocol. All probes

were prepared according to the manufacturers (Roche Molecular Biochemicals) protocols: Digoxigenin (DIG) labeled antisense transcripts was prepared from purified PCR products using a DIG RNA labeling kit with T3 RNA polymerase. DIG labeled PCR was prepared by using a DIG PCR labeling kit. 3"-DIG-tailed oligo ssDNA antisense probes, containing DIG-dUTP and dATP at an average tail length of 50 nucleotides were prepared from 100pmole oligonucleotides with the DIG Oligonucleotide Labeling Kit.

[0319] CLONE-SEQUENCING: PCR products were inserted into pGEM-T (Promega) or pTZ57 (MBI Fermentas), transformed into competent JM109 E. coli (Promega) and sown on LB-Amp plates with IPTG/Xgal. White and light-blue colonies were transferred to duplicate gridded plates, one of which was blotted onto a membrane (Biodyne Plus, Pall) for hybridization with DIG tailed oligo probes (according to instructions, Roche) corresponding to the expected GAM. Plasmid DNA from positive colonies was sequenced.

[0320] LIGATION-PCR DIAGNOSTIC METHOD: To further validate predicted GAM PCR product sequence derived from hemi-primers, a PCR based diagnostic technique was devised to amplify only those products containing also at least two

additional nucleotides of the non hemi-primer defined part of the predicted GAM RNA oligonucleotide. In essence, a diagnostic primer was designed so that its 3" end, which is the specificity determining side, was identical to the desired GAMRNA oligonucleotide, 2-10 nucleotides (typically 4-7, chosen for maximum specificity) further into its 3" end than the nucleotide stretch primed by the hemi-primer. The hemi-primer PCR product was first ligated into a T-cloning vector (pTZ57/T or pGEM-T) as described herinabove. The ligation reaction mixture was used as template for the diagnostic PCR under strict annealing conditions with the new diagnostic primer in conjunction with a general plasmid-homologous primer, resulting in a distinct ~200 base-pair product. This PCR product can be directly sequenced, permitting the elucidation of the remaining nucleotides up to the 3" of the mature GAM RNA oligonucleotide adjacent to the 3" adapter. Alternatively, following analysis of the diagnostic PCR reaction on an agarose gel, positive ligation reactions (containing a band of the expected size) were transformed into E. coli. Using this same diagnostic technique and as an alternative to screening by Southern-blot colony-hybridization, transformed bacterial colonies were

screened by colony-PCR (Gussow,D. and Clackson,T, Nucleic Acids Res. 17: 4000 (1989)) prior to plasmid purification and sequencing.

[0321] Reference is now made to Fig. 22B which is a table summarizing laboratory results which validate the expression of 8 known human miRNA oligonucleotide s and 43 novel GAM oligonucleotides detected by the BIOINFORMATIC OLIGONUCLEOTIDE DETECTION ENGINE 100. The table gives additional information on the above GAM oligonucleotides and contains the following fields:: NUMBER: refer to the hybridization picture number of Fig.22A; NAME: indicate the known MIR name for the reference set or GAM's name as given by RosettaGenomices nomenclature method; SEQUENCE: 5' to 3' sequence of the mature , 'diced' oligonucleotide; SEQUENCED: '+' indicates a validation of the GAM RNA sequence by sequencing procedure as described hereinabove with reference to Fig.22A.

[0322] Reference is now made to Fig. 23A, which is a schematic representation of a novel human GR polynucleotide herein designated GR12731 (RosettaGenomics Ltd. nomenclature), located on chromosome 9, comprising 2 known human MIR genes - MIR24 and MIR23, and 2 novel GAM oligonucleotides, herein designated GAM22 and GAM116,

all marked by solid black boxes. Fig. 23A also schematically illustrates 6 non-GAM hairpin sequences, and one non-hairpin sequence, all marked by white boxes, and serving as negative controls. By "non-GAM hairpin sequences" is meant sequences of a similar length to known MIR PRECURSOR sequences, which form hairpin secondary folding pattern similar to MIR PRECURSOR hairpins, and yet which are assessed by the BIOINFORMATIC OLIGONUCLEOTIDE DETECTION ENGINE 100 not to be valid GAM PRECURSOR hairpins. It is appreciated that Fig. 23A is a simplified schematic representation, reflecting only the order in which the segments of interest appear relative to one another, and not a proportional distance between the segments.

[0323] Reference is now made to Fig. 23B, which is a schematic representation of secondary folding of each of the MIRs and GAMs of GR GR12731 MIR24, MIR23, GAM22 and GAM116, and of the negative control non-GAM hairpins, herein designated N2, N3, N116, N4, N6 and N7. N0 is a non-hairpin control, of a similar length to that of known MIR PRECURSOR hairpins. It is appreciated that the negative controls are situated adjacent to and in between real MIR genes and GAM predicted oligonucleotide and

demonstrates similar secondary folding patterns to that of known MIRs and GAMs.

[0324] Reference is now made to Fig. 23C, which is a picture of laboratory results of a PCR test upon a YM100 "ligation"-library, utilizing specific primer sets directly inside the boundaries of the hairpins. Due to the nature of the library the only PCR amplifiable products can result from RNaseIII type enzyme cleaved RNA, as expected for legitimate hairpin precursors presumed to be produced by DROSHA (Lee et al, Nature 425 415-419, 2003). Fig 23C demonstrates expression of hairpin precursors of known MIR genes - MIRhsa-23 and MIRhsa-24, and of novel bioinformatically detected GAM22 and GAM116 hairpins predicted bioinformatically by a system constructed and operative in accordance with a preferred embodiment of the present invention. Fig. 23C also shows that none of the 7 controls (6 hairpins designated N2, N3, N23, N4, N6 and N7 and 1 non-hairpin sequence designated N0) were expressed. N116 is a negative control sequence partially overlapping GAM116.

[0325] In the picture, test lanes including template are designated "+" and the control lane is designated "-". It is appreciated that for each of the tested hairpins, a clear PCR

band appears in the test ("+") lane, but not in the control ("-") lane.

[0326] Figs. 23A through 23C, when taken together validate the efficacy of the bioinformatic oligonucleotide detection engine in: (a) detecting known MIR genes; (b) detecting novel GAM PRECURSOR hairpins which are found adjacent to these MIR genes, and which despite exhaustive prior biological efforts and bioinformatic detection efforts, went undetected; (c) discerning between GAM (or MIR) PRECURSOR hairpins, and non-GAM hairpins.

[0327] It is appreciated that the ability to discern GAM-hairpins from non-GAM-hairpins is very significant in detecting GAM oligonucleotide since hairpins in general are highly abundant in the genome. Other MIR prediction programs have not been able to address this challenge successfully.

[0328] Reference is now made to Fig. 24A which is an annotated sequence of an EST comprising a novel GAM oligonucleotides detected by the oligonucleotide detection system of the present invention. Fig. 24A shows the nucleotide sequence of a known human non-protein coding EST (Expressed Sequence Tag), identified as EST72223. The EST72223 clone obtained from TIGR database (Kirkness and Kerlavage, 1997) was sequenced to yield the

above 705bp transcript with a polyadenyl tail. It is appreciated that the sequence of this EST comprises sequences of one known miRNA oligonucleotide , identified as hsa-MIR98, and of one novel GAM oligonucleotide referred to here as GAM25, detected by the BIOINFORMATIC OLIGONUCLEOTIDE DETECTION ENGINE 100 (Fig. 9) of the present invention.

[0329] The sequences of the precursors of the known MIR98 and of the predicted GAM25 are precursor in bold, the sequences of the established miRNA 98 and of the predicted miRNA-like oligonucleotide GAM25 are underlined.

[0330] Reference is now made to Figs. 24B, 24C and 24D that are pictures of laboratory results, which when taken together demonstrate laboratory confirmation of expression of the bioinformatically detected novel oligonucleotide of Fig. 24A. In two parallel experiments, an enzymatically synthesized capped, EST72223 RNA transcript, was incubated with Hela S100 lysate for 0 minutes, 4 hours and 24 hours. RNA was subsequently harvested, run on a denaturing polyacrylamide gel, and reacted with a 102nt and a 145nt antisense MIR98 and GAM25 precursor transcript probes respectively. The Northern blot results of these experiments demonstrated processing of EST72223 RNA

by Hela lysate (lanes 2-4, in 24B and 24C), into ~80bp and ~22bp segments, which reacted with the MIR98 precursor probe (24B), and into ~100bp and ~24bp segments, which reacted with the GAM25 precursor probe (24C). These results demonstrate the processing of EST72223 by Hela lysate into MIR98 precursor and GAM25 precursor. It is also appreciated from Fig. 24C (lane 1) that Hela lysate itself reacted with the GAM25 precursor probe, in a number of bands, including a ~100bp band, indicating that GAM25-precursor is endogenously expressed in Hela cells. The presence of additional bands, higher than 100bp in lanes 5-9 probably corresponds to the presence of nucleotide sequences in Hela lysate, which contain the GAM25 sequence.

[0331] In addition, in order to demonstrate the kinetics and specificity of the processing of MIR98 and GAM25 precursors into their respective mature, 'diced' segments, transcripts of MIR98 and of the bioinformatically predicted GAM25 precursors were similarly incubated with Hela S100 lysate, for 0 minutes, 30 minutes, 1 hour and 24 hours, and for 24 hours with the addition of EDTA, added to inhibit Dicer activity, following which RNA was harvested, run on a polyacrylamide gel and reacted with

MIR98 and GAM25 precursor probes. Capped transcripts were prepared for in-vitro RNA cleavage assays with T7 RNA polymerase including a m7G(5')ppp(5')G-capping reaction the Message Machine kit (Ambion). Purified PCR products were used as template for the reaction. These were amplified for each assay with specific primers containing a T7 promoter at the 5" end and a T3 RNA polymerase promoter at the 3"end. Capped RNA transcripts were incubated at 30C in supplemented, dialysis concentrated, Hela S100 cytoplasmic extract (4C Biotech, Seneffe, Belgium). The Hela S100 was supplemented by dialysis to a final concentration of 20mM Hepes, 100mM KCl, 2.5mM MgCl2 , 0.5mM DTT, 20% glycerol and protease inhibitor cocktail tablets (Complete mini Roche Molecular Biochemicals). After addition of all components, final concentrations were 100mM capped target RNA, 2mM ATP, 0.2mM GTP, 500U/ml RNasin, 25g/ml creatine kinase, 25mM creatine phosphate, 2.5mM DTT and 50% S100 extract. Proteinase K, used to enhance Dicer activity (Zhang et al., EMBO J. 21, 5875-5885 (2002)) was dissolved in 50mM Tris-HCl pH 8, 5mM CaCl2, and 50% glycerol, was added to a final concentration of 0.6 mg/ml. Cleavage reactions were stopped by the addition of 8 volumes of proteinase K

buffer (200Mm Tris-Hcl, pH 7.5, 25m M EDTA, 300mM NaCl, and 2% SDS) and incubated at 65C for 15min at different time points (0, 0.5, 1, 4, 24h) and subjected to phenol/chloroform extraction. Pellets were dissolved in water and kept frozen. Samples were analyzed on a segmented half 6%, half 13% polyacrylamide 1XTBE-7M Urea gel.

[0332] The Northern blot results of these experiments demonstrated an accumulation of a ~22bp segment which reacted with the MIR98 precursor probe, and of a ~24bp segment which reacted with the GAM25 precursor probe, over time (lanes 5-8). Absence of these segments when incubated with EDTA (lane 9), which is known to inhibit Dicer enzyme (Zhang et al., 2002), supports the notion that the processing of MIR98 and GAM25 precursors into their ' diced' segments is mediated by Dicer enzyme, found in Hela lysate. The molecular sizes of EST72223, MIR-98 and GAM25 and their corresponding precursors are indicated by arrows.

[0333] Fig. 24D present Northern blot results of same above experiments with GAM25 probe (24nt). The results clearly demonstrated the accumulation of mature GAM25 oligonucleotide after 24 h.

[0334] To validate the identity of the band shown by the lower arrow in figs. 24C and 24D, a RNA band parallel to a marker of 24 base was excised from the gel and cloned as in Elbashir et al (2001) and sequenced. 90 clones corresponded to the sequence of mature GAM25 oligonucleotide ,three corresponded to GAM25* (the opposite arm of the hairpin with a 1-3 nucleotide 3" overhang) and two to the hairpin-loop.

[0335] GAM25 was also validated endogenously by sequencing from both sides from a HeLa YM100 total-RNA "ligation" libraries, utilizing hemispecific primers as described in Fig. 22.

[0336] Taken together, these results validate the presence and processing of a novel MIR-like oligonucleotide, GAM25, which was predicted bioinformatically. The processing of this novel GAM oligonucleotide product, by Hela lysate from EST72223, through its precursor, to its final form was similar to that observed for known miRNA oligonucleotide, MIR98.

[0337] Transcript products were 705nt (EST72223), 102nt (MIR98 precursor), 125nt (GAM25 precursor) long. EST72223 was PCR amplified with T7-EST 72223 forward primer: 5"-TAATACGACTCACTATAGGCCCTTATTAGAGGATTCTGC

T-3" and T3-EST72223 reverse primer:"-AATTAACCCTCACTAAAGGTTTTTTTTTCCTGAGACAGAGT-3".MIR98 was PCR amplified using EST72223 as a template with T7MIR98 forward primer: 5-"TAATACGACTCACTATAGGGTGAGGTAGTAAGTTGTATTGTT-3"and T3MIR98 reverse primer: 5"-AATTAACCCTCACTAAAGGGAAAGTAGTAAGTTGTATAGTT-3".GAM25 was PCR amplified using EST72223 as a template with GAM25 forward primer: 5"-GAGGCAGGAGAATTGCTTGA-3" and T3-EST72223 reverse primer:5"-AATTAACCCTCACTAAAGGCCTGAGACAGAGTCTTGCTC-3".

[0338] It is appreciated that the data presented in Figs. 24A, 24B,24C and 24D when taken together validate the function of the bioinformatic oligonucleotide detection engine 100 of Fig. 9. Fig. 24A shows a novel GAM oligonucleotide bioinformatically detected by the BIOINFORMATIC OLIGONUCLEOTIDE DETECTION ENGINE 100, and Figs. 24C and 24D show laboratory confirmation of the expression of this novel oligonucleotide. This is in accord with the engine training and validation methodology described hereinabove with reference to Fig. 10.

DETAILED DESCRIPTION OF LARGE TABLES

[0339] Table 1 comprises data relating the SEQ ID NO of GAM RNA oligonucleotides of the present invention to their corresponding GAM NAME, and contains the following fields: GAM SEQ-ID: GAM SEQ ID NO, as in the Sequence Listing; GAM NAME: Rosetta Genomics Ltd. nomenclature (see below); GAM RNA SEQUENCE: Sequence (5` to 3`) of the mature, `diced` GAM RNA; GAM POS: Dicer cut location (see below); and

[0340] Table 2 comprises detailed textual description according to the description of Fig.8 of each of a plurality of novel GAM oligonucleotide of the present invention, and contains the following fields: GAM NAME: Rosetta Genomics Ltd. nomenclature (see below);PRECUR SEQ-ID:GAM precursor Seq-ID, as in the Sequence Listing; PRECURSOR SEQUENCE: Sequence (5` to 3`) of the GAM precursor ; GAM DESCRIPTION: Detailed description of GAM oligonucleotide with reference to Fig.8; and

[0341] Table 3 comprises data relating to the source and location of novel GAM oligonucleotides of the present invention, and contains the following fields: GAM NAME: Rosetta Genomics Ltd. nomenclature (see below); PRECUR SEQ-ID: GAM precursor SEQ ID NO, as in the Sequence Listing; OR-

GANISM: Abbreviated (hsa = Homo sapiens); CHR: Chromosome encoding the GAM oligonucleotide; STRAND: Orientation on the chromosome, '+' for the plus strand,'-' for the minus strand; CHR-START OFFSET Start offset of GAM precursor sequence on the chromosome; CHR-END OFFSET: End offset of GAM precursor sequence on the chromosome; SOURCE_REF-ID: Accession number of source sequence; and

[0342] Table 4 comprises data relating to GAM precursors of novel GAM oligonucleotides of the present invention, and contains the following fields: GAM NAME: Rosetta Genomics Ltd. nomenclature (see below); PRECUR SEQ-ID: GAM precursor Seq-ID, as in the Sequence Listing; PRECURSOR-SEQUENCE: Sequence (5` to 3`) of the GAM precursor; GAM FOLDED PRECURSOR RNA: Schematic representation of the GAM folded precursor, beginning 5` end (beginning of upper row) to 3` end (beginning of lower row), where the hairpin loop is positioned at the right part of the draw; and

[0343] Table 5 comprises data relating to GAM oligonucleotides of the present invention, and contains the following fields: GAM NAME: Rosetta Genomics Ltd. nomenclature (see below); GAM RNA SEQUENCE: Sequence (5` to 3`) of the ma-

ture, `diced` GAM RNA ; PRECUR SEQ-ID : GAM precursor Seq-ID, as in the Sequence Listing; SOURCE_REF-ID: accession number of the source sequence; GAM POS: Dicer cut location (see below); and

[0344] Table 6 comprises data relating SEQ ID NO of the GAM target gene binding site sequence to TARGET gene name and target binding site sequence, and contains the following fields: TARGET BINDING SITE SEQ-ID: Target binding site SEQ ID NO, as in the Sequence Listing; TARGET: GAM target gene name; TARGET BINDING SITE SEQUENCE: Nucleotide sequence (5` to 3`) of the target binding site; and

[0345] Table 7 comprises data relating to target genes and binding sites of GAM oligonucleotides of the present invention, and contains the following fields: GAM NAME: Rosetta Genomics Ltd. nomenclature (see below); GAM RNA SEQUENCE: Sequence (5` to 3`) of the mature, `diced` GAM RNA; TARGET: GAM target gene name; TARGET REF-ID: Target accession number (GenBank); UTR: Untranslated region of binding site/s (3` or 5`); TARGET BS-SEQ: Nucleotide sequence (5` to 3`) of the target binding site; BINDING-SITE-DRAW: Schematic representation of the binding site, upper row represent 5` to 3` se-

quence of the GAM RNA, lower row represent 3` to 5` sequence of the target binding site; GAM POS: Dicer cut location (see below); and

[0346] Table 8 comprises data relating to functions and utilities of novel GAM oligonucleotides of the present invention, and contains the following fields: GAM NAME: Rosetta Genomics Ltd. nomenclature (see below);TARGET: GAM target gene name; GAM RNA SEQUENCE: Sequence (5` to 3`) of the mature, `diced` GAM RNA; GAM FUNCTION: Description of the GAM functions and utilities; GAM POS: Dicer cut location (see below);TAR DIS: Target Disease Relation Group (see below); and

[0347] Table 9 comprises data of GAM target gene function references- Bibliography and contains the following fields: GAM NAME: Rosetta Genomics Ltd. nomenclature (see below); GAM RNA SEQUENCE: Sequence (5` to 3`) of the mature, `diced` GAM RNA; TARGET: GAM target gene name; REFERENCES: list of references relating to the target gene; GAM POS: Dicer cut location (see below); and

[0348] Table 10 comprises data relating to novel GR (Genomic Record) polynucleotides of the present invention, and contains the following fields: GR NAME: Rosetta Genomics Ltd. nomenclature (see below); GR DESCRIPTION: Detailed

description of a GR polynucleotide cluster, with reference to Fig.16; and

[0349] Table 11 comprises data relating to Diseases that GAM oligonucleotides are predicted to regulate the disease-associated genes. Each row is referred to a specific disease, and list the GAM target genes related to the disease. The first row is a summary of ALL target genes associated in all diseases containing in the present invention. The table contains the following fields: ROW#: index of the row number; DISEASE NAME: name of the disease; TARGET GENES ASSOCIATED WITH DISEASE: list of GAM target genes that are associated with the specified disease; and

[0350] The following conventions and abbreviations are used in the tables: The nucleotide 'U' is represented as 'T' in the tables, and

[0351] GAM NAME or GR NAME are names for nucleotide sequences of the present invention given by RosettaGenomics Ltd. nomenclature method. All GAMs/GRs are designated by GAMx/GRx where x is a unique ID.

[0352] SOURCE REF-ID: The accession number of expressed sequences on which novel oligonucleotides were detected. The sequences are taken from the following published databases: (1) TIGR- "Tentative Human Consensus" (THC)

(2) EST database-UNIGENE, NCBI.

[0353] GAM POS is a position of the GAM RNA on the GAM PRECURSOR RNA sequence. This position is the Dicer cut location, 'A' indicates a probable Dicer cut location, 'B' indicates an alternative Dicer cut location.

[0354] TAR DIS (Target Disease Relation Group) 'A' indicates if the target gene is known to have a specific causative relation to a specific known disease, based on the OMIM database (Hamosh et al, 2002). It is appreciated that this is a partial classification emphasizing genes which are associated with "single gene" diseases etc. All genes of the present invention ARE associated with various diseases, although not all are in 'A' status.

[0355] All genomic sequences of the present invention as well as their chromosomal location and strand orientation are derived from sequences records of NCBI, Build33 database (April, 2003).

[0356] It is appreciated by persons skilled in the art that the present invention is not limited by what has been particularly shown and described hereinabove. Rather the scope of the present invention includes both combinations and subcombinations of the various features described hereinabove as well as variations and modifications which

would occur to persons skilled in the art upon reading the specifications and which are not in the prior art.